

T350x

## SEQUENCE LISTING

## (1) GENERAL INFORMATION:

(i) APPLICANT: Garner, Ian  
Dalrymple, Michael A  
Prunkard, Donna E  
Foster, Donald C

(ii) TITLE OF INVENTION: Production of Fibrinogen in Transgenic  
Animals

(iii) NUMBER OF SEQUENCES: 27

## (iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: ZymoGenetics, Inc.  
(B) STREET: 4225 Roosevelt Way, N.E.  
(C) CITY: Seattle  
(D) STATE: WA  
(E) COUNTRY: USA  
(F) ZIP: 98105

## (v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk  
(B) COMPUTER: IBM PC compatible  
(C) OPERATING SYSTEM: PC-DOS/MS-DOS  
(D) SOFTWARE: PatentIn Release #1.0, Version #1.25

## (vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER:  
(B) FILING DATE:  
(C) CLASSIFICATION:

## (viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: Parker, Gary E  
(B) REGISTRATION NUMBER: 31-648  
(C) REFERENCE/DOCKET NUMBER: 93-15

## (ix) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: 206-547-8080 ext 322  
(B) TELEFAX: 206-548-2329

35

## (2) INFORMATION FOR SEQ ID NO:1:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 5943 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA (genomic)

## (vii) IMMEDIATE SOURCE:

- (B) CLONE: Human Fibrinogen A-alpha chain

## (ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: join(31..84, 1154..1279, 1739..1922, 3055..3200, 3786..5210)

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

GTCTAGGAGC CAGCCCCACC CTTAGAAAAG ATG TTT TCC ATG AGG ATC GTC TGC	54
Met Phe Ser Met Arg Ile Val Cys	
1 5	
CTA GTT CTA AGT GTG GTG GGC ACA GCA TGG GTATGGCCCT TTTCATTTTT	104
Leu Val Leu Ser Val Val Gly Thr Ala Trp	
10 15	
TCTTCTTGCT TTCTCTCTGG TGTTTATTCC ACAAAGAGCC TGGAGGTCAG AGTCTACCTG	164
CTCTATGTCC TGACACACTC TTAGCTTTAT GACCCAGGC CTGGGAGGAA ATTCCTGGG	224
TGGGCTTGAC ACCTCAAGAA TACAGGGTAA TATGACACCA AGAGGAAGAT CTTAGATGGA	284
TGAGAGTGTA CAACTACAAG GGAAACTTTA GCATCTGTCA TTCAGTCTTA CCACATTTTG	344
TTTTGTTTTG TTTTAAAAAG GGCAAGAATT ATTTGCCATC CTTGTACCTA TAAAGCCTTG	404
GTGCATTATA ATGCTAGTTA ATGGAATAAA ACATTTTATG GTAAGATTTG TTTTCTTTAG	464
TTATTAATTT CTTGCTACTT GTCCATAATA AGCAGAACTT TTAGTGTTAG TACAGTTTTG	524
CTGAAAGGTT ATTGTTGTGT TTGTCAAGAC AGAAGAAAAA GCAAACGAAT TATCTTTGGA	584

36

AATATCTTTG CAGTATCAGA AGAGATTAGT TAGTAAGGCA ATACGCTTTT CCGCAGTAAT	644
GGTATTCTTT TAAATTATGA ATCCATCTCT AAAGGTTACA TAGAACTTG AAGGAGAGAG	704
GAACATTCAG TTAAGATAGT CTAGGTTTTT CTACTGAAGC AGCAATTACA GGAGAAAGAG	764
CTCTACAGTA GTTTTCAACT TTCTGTCTGC AGTCATTAGT AAAAATGAAA AGGTAAAATT	824
TAACTGATTT TATAGATTCA AATAATTTTC CTTTATAGGAT GGATTCTTTA AAACTCCTAA	884
TATTTATCAA ATGCTTATTT AAGTGTCAACA CACAGTTAAG AAATTTGTAC ACCTTGTCTC	944
CTTTAATTCT CATAACAACT CCATAAAATG GGTCCTAGGA TTTCCATTTG AAGATAAGAA	1004
ACCTGAAGCT TGCCGAAGCC CTGTGTCTGC TCTCCTTAAT CTCTGTGAGA GTGCCATCTC	1064
TTCTGGGGA CTTGTAGGCA TGCCACTGTC TCCTCTTCTG GCTAACATTG CTGTTGCTCT	1124
CTTTTGTGTA TGTGAATGAA TCTTTAAAG ACT GCA GAT AGT GGT GAA GGT GAC	1177
Thr Ala Asp Ser Gly Glu Gly Asp	
20 25	
TTT CTA GCT GAA GGA GGA GGC GTG CGT GGC CCA AGG GTT GTG GAA AGA	1225
Phe Leu Ala Glu Gly Gly Gly Val Arg Gly Pro Arg Val Val Glu Arg	
30 35 40	
CAT CAA TCT GCC TGC AAA GAT TCA GAC TGG CCC TTC TGC TCT GAT GAA	1273
His Gln Ser Ala Cys Lys Asp Ser Asp Trp Pro Phe Cys Ser Asp Glu	
45 50 55	
GAC TGG GTAAGCAGTC AGCGGGGGAA GCAGGAGATT CCTTCCCTCT GATGCTAGAG	1329
Asp Trp	
60	
GGGCTCACAG GCTGACCTGA TTGGTCCCAG AAACTTTTTT AAATAGAAAA TAATTGAATA	1389
GTTACCTACA TAGCAAATAA AGAAAAGGAA CCTACTCCCA AGAGCACTGT TTATTACCT	1449
CCCCAACTCT GGATCATTAG TGGGTGAACA GACAGGATTT CAGTTGCATG CTCAGGCAAA	1509
ACCAGGCTCC TGAGTATTGT GGCCTCAATT TCCTGGCACC TATTTATGGC TAAGTGGACC	1569
CTCATTCCAG AGTTTCTCTG CGACCTCTAA CTAGTCCTCT TACCTACTTT TAAGCCAAC	1629

TATCTGGAAG AGAAAGGGTA GGAAGAAATG GGGGCTGCAT GGAAACATGC AAAATTATTC	1689
TGAATCTGAG AGATAGATCC TTAAGTAAAT TTTCTCCCTT CACTTTCAG AAC TAC Asn Tyr	1744
AAA TGC CCT TCT GGC TGC AGG ATG AAA GGG TTG ATT GAT GAA GTC AAT Lys Cys Pro Ser Gly Cys Arg Met Lys Gly Leu Ile Asp Glu Val Asn 65 70 75	1792
CAA GAT TTT ACA AAC AGA ATA AAT AAG CTC AAA AAT TCA CTA TTT GAA Gln Asp Phe Thr Asn Arg Ile Asn Lys Leu Lys Asn Ser Leu Phe Glu 80 85 90	1840
TAT CAG AAG AAC AAT AAG GAT TCT CAT TCG TTG ACC ACT AAT ATA ATG Tyr Gln Lys Asn Asn Lys Asp Ser His Ser Leu Thr Thr Asn Ile Met 95 100 105 110	1888
GAA ATT TTG AGA GGC GAT TTT TCC TCA GCC AAT A GTAAGTATTA Glu Ile Leu Arg Gly Asp Phe Ser Ser Ala Asn 115 120	1932
CATATTTACT TCTTTGACTT TATAACAGAA ACAACAAAA TCCTAAATAA ATATGATATC	1992
CGCTTATATC TATGACAATT TCATCCCAAA GTACTTAGTG TAGAAACACA TACCTTCATA	2052
ATATCCCTGA AAATTTTAAG AGGGAGCTTT TGTTTTCGTT ATTTTTTCAA AGTAAAAGAT	2112
GTAACTGAG ATTGTTTAAG GTCACAAAAT AAGTCAGAAT TTTGGATTAA AACAAGAATT	2172
TAAATGTGTT CTTTTCAACA GTATATACTG AAAGTAGGAT GGGTCAGACT CTTTGAGTTG	2232
ATATTTTTGT TTCTGCTTTG TAAAGGTGAA AACTGAGAGG TCAAGGAAC TGTTCAAAGA	2292
CACAGAGCTG GGAATTCAAC TCCAGACTC CACTGAGCTG ATTAGGTAGA TTTTAAATT	2352
TAAAATATAG GGTCAAGCTA CGTCATTCTC ACAGTCTACT CATTAGGGTT AGGAAACATT	2412
GCATTCACTC TGGGCATGGA CAGCGAGTCT AGGGAGTCCT CAGTTTCTCA AGTTTTGCTT	2472
TGCCTTTTTA CACCTTCACA AACACTTGAC ATTTAAAATC AGTGATGCCA AACTAGCTG	2532
GCAAGTGAGT GATCCTGTTG ACCCAAAACA GCTTAGGAAC CATTTCAAAT CTATAGAGTT	2592
AAAAAGAAAA GCTCATCAGT AAGAAAATCC AATATGTTCA AGTCCCTTGA TTAAGGATGT	2652

TATAAAATAA TTGAAATGCA ATCAAACCAA CTATTTTAAC TCCAAATTAC ACCTTTAAAA	2712
TTCCAAAGAA AGTTCTTCTT CTATATTCTT TTGGGATTAC TAATTGCTAT TAGGACATCT	2772
TAACTGGCAT TCATGGAAGG CTGCAGGGCA TAACATTATC CAAAAGTCAA ATGCCCCATA	2832
GGTTTTGAAC TCACAGATTA AACTGTAACC AAAATAAAAT TAGGCATATT TACAAGCTAG	2892
TTTCTTTCTT TCTTTTTTCT CTTTCTTTCT TTCTTTCTTT CTTTCTTTCT TTCTTTCTTT	2952
CTTTCTTTCT TTCTCCTTCC TTCCTTTCTT CCTTTCTTTT TTGCTGGCAA TTACAGACAA	3012
ATCACTCAGC AGCTACTTCA ATAACCATAT TTTCGATTTC AG AC CGT GAT AAT	3065
	Asn Arg Asp Asn
	125
ACC TAC AAC CGA GTG TCA GAG GAT CTG AGA AGC AGA ATT GAA GTC CTG	3113
Thr Tyr Asn Arg Val Ser Glu Asp Leu Arg Ser Arg Ile Glu Val Leu	
	130 135 140
AAG CGC AAA GTC ATA GAA AAA GTA CAG CAT ATC CAG CTT CTG CAG AAA	3161
Lys Arg Lys Val Ile Glu Lys Val Gln His Ile Gln Leu Leu Gln Lys	
	145 150 155
AAT GTT AGA GCT CAG TTG GTT GAT ATG AAA CGA CTG GAG GTAAGTATGT	3210
Asn Val Arg Ala Gln Leu Val Asp Met Lys Arg Leu Glu	
	160 165 170
GGCTGTGGTC CCGAGTGTCC TTGTTTTTGA GTAGAGGGAA AAGGAAGGCG ATAGTTATGC	3270
ACTGAGTGTC TACTATATGC AGAGAAAAGT GTTATATCCA TCATCTACCT AAAAGTAGGT	3330
ATTATTTTCC TCACTCCACA GTTGAAGAAA AAAAAATTCA GAGATATTAA GTAAATTTTC	3390
CAACGTACAT AGATAGTAAT TCAAAGCAAT GTTCAGTCCC TGTCTATTCC AAGCCATTAC	3450
ATCACCACAC CTCTGAGCCC TCAGCCTGAG TTCACCAAGG ATCATTTAAT TAGCGTTTCC	3510
TTTGAGAGGG AATAGCACCT TACTCTTGAT CCATTCTGAG GCTAAGATGA ATTAAACAGC	3570
ATCCATTGCT TATCCTGGCT AGCCCTGCAA TACCCAACAT CTCTTCCACT GAGGGTGCTC	3630
GATAGGCAGA AAACAGAGAA TATTAAGTGG TAGGTCTCCG AGTCAAAAAA AATGAAACCA	3690

GTTCAGAA GGAAATTA CTACCAGGAA CTCAATAGAC GTAGTTTATG TATTTGTATC	3750
TACATTTTCT CTTTATTTTT CTCCCCTCTC TCTAG GTG GAC ATT GAT ATT AAG	3803
Val Asp Ile Asp Ile Lys	175
ATC CGA TCT TGT CGA GGG TCA TGC AGT AGG GCT TTA GCT CGT GAA GTA	3851
Ile Arg Ser Cys Arg Gly Ser Cys Ser Arg Ala Leu Ala Arg Glu Val	180 185 190
GAT CTG AAG GAC TAT GAA GAT CAG CAG AAG CAA CTT GAA CAG GTC ATT	3899
Asp Leu Lys Asp Tyr Glu Asp Gln Gln Lys Gln Leu Glu Gln Val Ile	195 200 205
GCC AAA GAC TTA CTT CCC TCT AGA GAT AGG CAA CAC TTA CCA CTG ATA	3947
Ala Lys Asp Leu Leu Pro Ser Arg Asp Arg Gln His Leu Pro Leu Ile	210 215 220
AAA ATG AAA CCA GTT CCA GAC TTG GTT CCC GGA AAT TTT AAG AGC CAG	3995
Lys Met Lys Pro Val Pro Asp Leu Val Pro Gly Asn Phe Lys Ser Gln	225 230 235 240
CTT CAG AAG GTA CCC CCA GAG TGG AAG GCA TTA ACA GAC ATG CCG CAG	4043
Leu Gln Lys Val Pro Pro Glu Trp Lys Ala Leu Thr Asp Met Pro Gln	245 250 255
ATG AGA ATG GAG TTA GAG AGA CCT GGT GGA AAT GAG ATT ACT CGA GGA	4091
Met Arg Met Glu Leu Glu Arg Pro Gly Gly Asn Glu Ile Thr Arg Gly	260 265 270
GGC TCC ACC TCT TAT GGA ACC GGA TCA GAG ACG GAA AGC CCC AGG AAC	4139
Gly Ser Thr Ser Tyr Gly Thr Gly Ser Glu Thr Glu Ser Pro Arg Asn	275 280 285
CCT AGC AGT GCT GGA AGC TGG AAC TCT GGG AGC TCT GGA CCT GGA AGT	4187
Pro Ser Ser Ala Gly Ser Trp Asn Ser Gly Ser Ser Gly Pro Gly Ser	290 295 300
ACT GGA AAC CGA AAC CCT GGG AGC TCT GGG ACT GGA GGG ACT GCA ACC	4235
Thr Gly Asn Arg Asn Pro Gly Ser Ser Gly Thr Gly Gly Thr Ala Thr	305 310 315 320
TGG AAA CCT GGG AGC TCT GGA CCT GGA AGT GCT GGA AGC TGG AAC TCT	4283
Trp Lys Pro Gly Ser Ser Gly Pro Gly Ser Ala Gly Ser Trp Asn Ser	325 330 335

GGG AGC TCT GGA ACT GGA AGT ACT GGA AAC CAA AAC CCT GGG AGC CCT	4331
Gly Ser Ser Gly Thr Gly Ser Thr Gly Asn Gln Asn Pro Gly Ser Pro	
340 345 350	
AGA CCT GGT AGT ACC GGA ACC TGG AAT CCT GGC AGC TCT GAA CGC GGA	4379
Arg Pro Gly Ser Thr Gly Thr Trp Asn Pro Gly Ser Ser Glu Arg Gly	
355 360 365	
AGT GCT GGG CAC TGG ACC TCT GAG AGC TCT GTA TCT GGT AGT ACT GGA	4427
Ser Ala Gly His Trp Thr Ser Glu Ser Ser Val Ser Gly Ser Thr Gly	
370 375 380	
CAA TGG CAC TCT GAA TCT GGA AGT TTT AGG CCA GAT AGC CCA GGC TCT	4475
Gln Trp His Ser Glu Ser Gly Ser Phe Arg Pro Asp Ser Pro Gly Ser	
385 390 395 400	
GGG AAC GCG AGG CCT AAC AAC CCA GAC TGG GGC ACA TTT GAA GAG GTG	4523
Gly Asn Ala Arg Pro Asn Asn Pro Asp Trp Gly Thr Phe Glu Glu Val	
405 410 415	
TCA GGA AAT GTA AGT CCA GGG ACA AGG AGA GAG TAC CAC ACA GAA AAA	4571
Ser Gly Asn Val Ser Pro Gly Thr Arg Arg Glu Tyr His Thr Glu Lys	
420 425 430	
CTG GTC ACT TCT AAA GGA GAT AAA GAG CTC AGG ACT GGT AAA GAG AAG	4619
Leu Val Thr Ser Lys Gly Asp Lys Glu Leu Arg Thr Gly Lys Glu Lys	
435 440 445	
GTC ACC TCT GGT AGC ACA ACC ACC ACG CGT CGT TCA TGC TCT AAA ACC	4667
Val Thr Ser Gly Ser Thr Thr Thr Thr Arg Arg Ser Cys Ser Lys Thr	
450 455 460	
GTT ACT AAG ACT GTT ATT GGT CCT GAT GGT CAC AAA GAA GTT ACC AAA	4715
Val Thr Lys Thr Val Ile Gly Pro Asp Gly His Lys Glu Val Thr Lys	
465 470 475 480	
GAA GTG GTG ACC TCC GAA GAT GGT TCT GAC TGT CCC GAG GCA ATG GAT	4763
Glu Val Val Thr Ser Glu Asp Gly Ser Asp Cys Pro Glu Ala Met Asp	
485 490 495	
TTA GGC ACA TTG TCT GGC ATA GGT ACT CTG GAT GGG TTC CGC CAT AGG	4811
Leu Gly Thr Leu Ser Gly Ile Gly Thr Leu Asp Gly Phe Arg His Arg	
500 505 510	

CAC CCT GAT GAA GCT GCC TTC TTC GAC ACT GCC TCA ACT GGA AAA ACA His Pro Asp Glu Ala Ala Phe Phe Asp Thr Ala Ser Thr Gly Lys Thr 515 520 525	4859
TTC CCA GGT TTC TTC TCA CCT ATG TTA GGA GAG TTT GTC AGT GAG ACT Phe Pro Gly Phe Phe Ser Pro Met Leu Gly Glu Phe Val Ser Glu Thr 530 535 540	4907
GAG TCT AGG GGC TCA GAA TCT GGC ATC TTC ACA AAT ACA AAG GAA TCC Glu Ser Arg Gly Ser Glu Ser Gly Ile Phe Thr Asn Thr Lys Glu Ser 545 550 555 560	4955
AGT TCT CAT CAC CCT GGG ATA GCT GAA TTC CCT TCC CGT GGT AAA TCT Ser Ser His His Pro Gly Ile Ala Glu Phe Pro Ser Arg Gly Lys Ser 565 570 575	5003
TCA AGT TAC AGC AAA CAA TTT ACT AGT AGC ACG AGT TAC AAC AGA GGA Ser Ser Tyr Ser Lys Gln Phe Thr Ser Ser Thr Ser Tyr Asn Arg Gly 580 585 590	5051
GAC TCC ACA TTT GAA AGC AAG AGC TAT AAA ATG GCA GAT GAG GCC GGA Asp Ser Thr Phe Glu Ser Lys Ser Tyr Lys Met Ala Asp Glu Ala Gly 595 600 605	5099
AGT GAA GCC GAT CAT GAA GGA ACA CAT AGC ACC AAG AGA GGC CAT GCT Ser Glu Ala Asp His Glu Gly Thr His Ser Thr Lys Arg Gly His Ala 610 615 620	5147
AAA TCT CGC CCT GTC AGA GGT ATC CAC ACT TCT CCT TTG GGG AAG CCT Lys Ser Arg Pro Val Arg Gly Ile His Thr Ser Pro Leu Gly Lys Pro 625 630 635 640	5195
TCC CTG TCC CCC TAGACTAAGT TAAATATTTT TGCACAGTGT TCCCATGGCC Ser Leu Ser Pro 645	5247
CCTTGCATTT CTTCTTAAC TCTCTGTTAC ACGTCATTGA AACTACACTT TTTTGGTCTG	5307
TTTTTGTGCT AGACTGTAAG TTCCTTGGGG GCAGGGCCTT TGTCTGTCTC ATCTCTGTAT	5367
TCCCAAATGC CTAACAGTAC AGAGCCATGA CTCAATAAAT ACATGTAAAA TGGATGAATG	5427
AATTCCTCTG AAACCTCTATT TGAGCTTATT TAGTCAAATT CTTTCACTAT TCAAAGTGTG	5487
TGCTATTAGA ATTGTCACCC AACTGATTAA TCACATTTTT AGTATGTGTC TCAGTTGACA	5547



TTTAGGTCAG GCTAAATACA AGTTGTGTTA GTATTAAGTG AGCTTAGCTA CCTGTACTGG 5607  
 TTACTIONGCTA TTAGTTTGTG CAAGTAAAAT TCCAAATACA TTTGAGGAAA ATCCCCTTTG 5667  
 CAATTTGTAG GTATAAATAA CCGCTTATTT GCATAAGTTC TATCCCACTG TAAGTGCATC 5727  
 CTTTCCCTAT GGAGGGAAGG AAAGGAGGAA GAAAGAAAGG AAGGGAAAGA AACAGTATTT 5787  
 GCCTTATTTA ATCTGAGCCG TGCCTATCTT TGTAAGTTA AATGAGAATA ACTTCTTCCA 5847  
 ACCAGCTTAA TTTTTTTTTT AGACTGTGAT GATGTCCTCC AAACACATCC TTCAGGTACC 5907  
 CAAAGTGGCA TTTTCAATAT CAAGCTATCC GGATCC 5943

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 644 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met	Phe	Ser	Met	Arg	Ile	Val	Cys	Leu	Val	Leu	Ser	Val	Val	Gly	Thr
1				5				10						15	
Ala	Trp	Thr	Ala	Asp	Ser	Gly	Glu	Gly	Asp	Phe	Leu	Ala	Glu	Gly	Gly
			20					25					30		
Gly	Val	Arg	Gly	Pro	Arg	Val	Val	Glu	Arg	His	Gln	Ser	Ala	Cys	Lys
		35				40					45				
Asp	Ser	Asp	Trp	Pro	Phe	Cys	Ser	Asp	Glu	Asp	Trp	Asn	Tyr	Lys	Cys
	50					55					60				
Pro	Ser	Gly	Cys	Arg	Met	Lys	Gly	Leu	Ile	Asp	Glu	Val	Asn	Gln	Asp
65					70				75					80	
Phe	Thr	Asn	Arg	Ile	Asn	Lys	Leu	Lys	Asn	Ser	Leu	Phe	Glu	Tyr	Gln
			85					90						95	

Lys Asn Asn Lys Asp Ser His Ser Leu Thr Thr Asn Ile Met Glu Ile  
 100 105 110  
 Leu Arg Gly Asp Phe Ser Ser Ala Asn Asn Arg Asp Asn Thr Tyr Asn  
 115 120 125  
 Arg Val Ser Glu Asp Leu Arg Ser Arg Ile Glu Val Leu Lys Arg Lys  
 130 135 140  
 Val Ile Glu Lys Val Gln His Ile Gln Leu Leu Gln Lys Asn Val Arg  
 145 150 155 160  
 Ala Gln Leu Val Asp Met Lys Arg Leu Glu Val Asp Ile Asp Ile Lys  
 165 170 175  
 Ile Arg Ser Cys Arg Gly Ser Cys Ser Arg Ala Leu Ala Arg Glu Val  
 180 185 190  
 Asp Leu Lys Asp Tyr Glu Asp Gln Gln Lys Gln Leu Glu Gln Val Ile  
 195 200 205  
 Ala Lys Asp Leu Leu Pro Ser Arg Asp Arg Gln His Leu Pro Leu Ile  
 210 215 220  
 Lys Met Lys Pro Val Pro Asp Leu Val Pro Gly Asn Phe Lys Ser Gln  
 225 230 235 240  
 Leu Gln Lys Val Pro Pro Glu Trp Lys Ala Leu Thr Asp Met Pro Gln  
 245 250 255  
 Met Arg Met Glu Leu Glu Arg Pro Gly Gly Asn Glu Ile Thr Arg Gly  
 260 265 270  
 Gly Ser Thr Ser Tyr Gly Thr Gly Ser Glu Thr Glu Ser Pro Arg Asn  
 275 280 285  
 Pro Ser Ser Ala Gly Ser Trp Asn Ser Gly Ser Ser Gly Pro Gly Ser  
 290 295 300  
 Thr Gly Asn Arg Asn Pro Gly Ser Ser Gly Thr Gly Gly Thr Ala Thr  
 305 310 315 320  
 Trp Lys Pro Gly Ser Ser Gly Pro Gly Ser Ala Gly Ser Trp Asn Ser  
 325 330 335

Gly Ser Ser Gly Thr Gly Ser Thr Gly Asn Gln Asn Pro Gly Ser Pro  
 340 345 350  
 Arg Pro Gly Ser Thr Gly Thr Trp Asn Pro Gly Ser Ser Glu Arg Gly  
 355 360 365  
 Ser Ala Gly His Trp Thr Ser Glu Ser Ser Val Ser Gly Ser Thr Gly  
 370 375 380  
 Gln Trp His Ser Glu Ser Gly Ser Phe Arg Pro Asp Ser Pro Gly Ser  
 385 390 395 400  
 Gly Asn Ala Arg Pro Asn Asn Pro Asp Trp Gly Thr Phe Glu Glu Val  
 405 410 415  
 Ser Gly Asn Val Ser Pro Gly Thr Arg Arg Glu Tyr His Thr Glu Lys  
 420 425 430  
 Leu Val Thr Ser Lys Gly Asp Lys Glu Leu Arg Thr Gly Lys Glu Lys  
 435 440 445  
 Val Thr Ser Gly Ser Thr Thr Thr Thr Arg Arg Ser Cys Ser Lys Thr  
 450 455 460  
 Val Thr Lys Thr Val Ile Gly Pro Asp Gly His Lys Glu Val Thr Lys  
 465 470 475 480  
 Glu Val Val Thr Ser Glu Asp Gly Ser Asp Cys Pro Glu Ala Met Asp  
 485 490 495  
 Leu Gly Thr Leu Ser Gly Ile Gly Thr Leu Asp Gly Phe Arg His Arg  
 500 505 510  
 His Pro Asp Glu Ala Ala Phe Phe Asp Thr Ala Ser Thr Gly Lys Thr  
 515 520 525  
 Phe Pro Gly Phe Phe Ser Pro Met Leu Gly Glu Phe Val Ser Glu Thr  
 530 535 540  
 Glu Ser Arg Gly Ser Glu Ser Gly Ile Phe Thr Asn Thr Lys Glu Ser  
 545 550 555 560  
 Ser Ser His His Pro Gly Ile Ala Glu Phe Pro Ser Arg Gly Lys Ser  
 565 570 575

Ser Ser Tyr Ser Lys Gln Phe Thr Ser Ser Thr Ser Tyr Asn Arg Gly  
                   580                                  585                                  590

Asp Ser Thr Phe Glu Ser Lys Ser Tyr Lys Met Ala Asp Glu Ala Gly  
                   595                                  600                                  605

Ser Glu Ala Asp His Glu Gly Thr His Ser Thr Lys Arg Gly His Ala  
                   610                                  615                                  620

Lys Ser Arg Pro Val Arg Gly Ile His Thr Ser Pro Leu Gly Lys Pro  
                   625                                  630                                  635                                  640

Ser Leu Ser Pro

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 8878 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vii) IMMEDIATE SOURCE:

- (B) CLONE: human fibrinogen B-beta chain

(ix) FEATURE:

- (A) NAME/KEY: misc\_RNA
- (B) LOCATION: 1..469

(ix) FEATURE:

- (A) NAME/KEY: exon
- (B) LOCATION: 470..583

(ix) FEATURE:

- (A) NAME/KEY: intron
- (B) LOCATION: 584..3257

(ix) FEATURE:

- (A) NAME/KEY: exon
- (B) LOCATION: 3258..3449

46

- (ix) FEATURE:  
    (A) NAME/KEY: intron  
    (B) LOCATION: 3450..3938
- (ix) FEATURE:  
    (A) NAME/KEY: exon  
    (B) LOCATION: 3939..4122
- (ix) FEATURE:  
    (A) NAME/KEY: intron  
    (B) LOCATION: 4123..5042
- (ix) FEATURE:  
    (A) NAME/KEY: exon  
    (B) LOCATION: 5043..5270
- (ix) FEATURE:  
    (A) NAME/KEY: intron  
    (B) LOCATION: 5271..5830
- (ix) FEATURE:  
    (A) NAME/KEY: exon  
    (B) LOCATION: 5831..5944
- (ix) FEATURE:  
    (A) NAME/KEY: intron  
    (B) LOCATION: 5945..6632
- (ix) FEATURE:  
    (A) NAME/KEY: exon  
    (B) LOCATION: 6633..6758
- (ix) FEATURE:  
    (A) NAME/KEY: intron  
    (B) LOCATION: 6759..6966
- (ix) FEATURE:  
    (A) NAME/KEY: exon  
    (B) LOCATION: 6967..7252
- (ix) FEATURE:  
    (A) NAME/KEY: intron  
    (B) LOCATION: 7253..7870

## (ix) FEATURE:

(A) NAME/KEY: exon  
(B) LOCATION: 7871..8102

## (ix) FEATURE:

(A) NAME/KEY: 3'UTR  
(B) LOCATION: 8103..8537

## (ix) FEATURE:

(A) NAME/KEY: misc\_RNA  
(B) LOCATION: 8538..8878

## (ix) FEATURE:

(A) NAME/KEY: CDS  
(B) LOCATION: join(470..583, 3258..3449, 3939..4122,  
5043..5270,  
5831..5944, 6633..6758, 6967..7252, 7871..8102)

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

GAATTCATGC CCCTTTTGAA ATAGACTTAT GTCATTGTCA GAAAACATAA GCATTTATGG	60
TATATCATTAT ATGAGTCACG ATTTTAGTGG TTGCCTTGTG AGTAGGTCAA ATTTACTAAG	120
CTTAGATTTG TTTTCTCACA TATTCTTTTCG GAGCTTGTGT AGTTTCCACA TTAATTTACC	180
AGAAACAAGA TACACACTCT CTTTGAGGAG TGCCCTAACT TCCCATCATT TTGTCCAATT	240
AAATGAATTG AAGAAATTTA ATGTTTCTAA ACTAGACCAA CAAAGAATAA TAGTTGTATG	300
ACAAGTAAAT AAGCTTTGCT GGGGAAGATGT TGCTTAAATG ATAAATGGT TCAGCCAACA	360
AGTGAACCAA AAATTAATA TTAATAAGG AAAGGTAACC ATTTCTGAAG TCATTCCTAG	420
CAGAGGACTC AGATATATAT AGGATTGAAG ATCTCTCAGT TAAGTCTAC ATG AAA	475
	Met Lys
	1
AGG ATG GTT TCT TGG AGC TTC CAC AAA CTT AAA ACC ATG AAA CAT CTA	523
Arg Met Val Ser Trp Ser Phe His Lys Leu Lys Thr Met Lys His Leu	
5 10 15	
TTA TTG CTA CTA TTG TGT GTT TTT CTA GTT AAG TCC CAA GGT GTC AAC	571
Leu Leu Leu Leu Leu Cys Val Phe Leu Val Lys Ser Gln Gly Val Asn	

20	25	30	
GAC AAT GAG GAG GTGAATTTT TAAAGCATT	TTATATTATT	AGTAGTATTA	623
Asp Asn Glu Glu			
35			
TTAATATAAG ATGTAACATA ATCATATTAT	GTGCTTATTT	TAATGAAATT	683
ATAGTTATGA AATGGAATTG TTAACCTCTG	ACTTATTGTA	TTTAAAGAAT	743
ATTTCTTATA TAAAAACAAA GTAATTTCTT	GTTTTCTAGT	TTATCACCTT	803
AGATGAGGAT GGCTTAGCTA ATGTAAGATG	TGTTTTTCTC	ACTTGCTATT	863
TGATTTTCAT TTAATTCTAG CAATACAGGA	TTACAATTAA	GAGGACAAGA	923
TCACAACTA TAAAATAATA AAAGAGCAGA	ATTTTAAGAT	AAAAGAACT	983
AGATTGTTCT TTGGTGAAGG AAGGTAATAT	ATATTGTTAC	TGAGATTACT	1043
ATTATAACTA AGCCTAAAAG CAAAATACAT	CAAGTGTAA	GATAGAAAAT	1103
TTTTTTCAGA TGAAAAGTTC AAATTAGAGT	TAGTGTGTAT	TGTTATTATT	1163
AAACACGGTT CAGTCTAATT TATTTATTTG	TAGAACAGTT	TGTCCTCAAC	1223
GCTGACTTAT TGCTGTTAAT TTGCAGTTAC	TAAAATACA	GAAATGCATT	1283
GATATTTAAG AAATTTAAAT TTTATCATCA	AACGTATCAT	GGCCAAATTT	1343
GCATAGTATC ATTAACTAG AAATAAGAAT	ACACAATAAT	ATTTAAATGA	1403
TTCGGATCAT TATTGAGTTT CAAGGGAAC	TGAGTGTTGT	ACTTATCAGA	1463
AAGAACATAT AGTTAATCTG GTTGTGTGTG	TAAAAACATA	TGGTTAATCT	1523
GGTTAATCAT ATTAGGTAAG AAAAATGTAA	AGAATGTGTA	AGACGAAATT	1583
ACTCTGCAAA GCACTTTCAC ATTTCTGCTT	ATCAACTAAA	CCTCACAGAG	1643
AGTTTAGGCT TTAAATGGA TTTTGATTAT	TCAACAAGTG	GCCTTCATAA	1703
TGTTTTTCTT TAAGTATATA CTTTCTTTAA	ATATTTTTTA	AAATTCCTT	1763
AAGCCAGACC ATCCATGCTA CCTCTCTAGT	GGCACTCTGA	AATAAAAAGA	1823
AAATAGTTTT			

CTCTGTTATA ATTGTATTTG TAATAAGCAG ATGAATCACA TTTCTTAAAA TTTGTTTTAG	1883
AGAGGGTAAG CTCTGACTAG GACCATGACT TCAATGTGAA ATATGTATAT ATCCTCCGAA	1943
TCTTTACATA TTAAGAATGT ATATAGTCAA CTGGTTAAAC AGGAAAATCT GGAACAGCCT	2003
GGCTGGGTTT TAATCTTAGC ACCATCCTAC TAAATGTTAA ATAATATTAT AATCTAATGA	2063
ATAAATGACA ATGCAATTCC AAATAGAGTT CATCTGATGA CTTCTAGACT CACAAAATTG	2123
CAAGAGAGCT CAGTTGTTGC TCAGTTGTTT CAAATCATGT CGTTTGTTAA TTTGTAATTA	2183
AGCTCCAAAG GATGTATAGC TACTGACAAA AAAAAAATG AGAATGTAGT TAATCCAAAT	2243
CAAACTTTC CTATTGCAAT GCGTATTTTC TGCTTCATTA TCCTTTAATA TAATATTTTA	2303
AGTTAGCAAG TAATTTTAAT TACAATGCAC AAGCCTTGAG AATTATTTTA AATATAAGAA	2363
AATCATAATG TTTGATAAAG AAATCATGTA AGAAATTTCA AGATAATGGT TTAACAAATA	2423
ATTTTGTTGA TAGAAGATAA GACTAAAAGT GAAATTCGAA GTGGAGAGGA CACTTAACT	2483
GTAGTACTTG TTATGTGTGA TTCCAGTAAA AATAGTAATG AGCACTTATT ATTGCCAAGT	2543
ACTGTTCTGA GGGTACCATA TGCAATAAGT TATTTAATCC TTACAATAAT CTTGTAAGGC	2603
AGATTCAAAC TATCATTACA CTTATTTTAC AGATGAGAAA ACTGGGGCAC AGATAAAGCA	2663
ACTTGCCCAA GGTCTCATAG CTGTAAGTCA ACCCTACGGT CAAGACCTAC AAGTAGCCGA	2723
GCTCCAGAGT ACATTATGAG GGTCAAAGAT TGTCTTATTA CAAATAAATT CCAAGTAGAA	2783
TCAACCTTTA ATAAGTCTTT AATGTCTCTT AAATATGTTT ATATAGGAGT CTAATCACCA	2843
ATTCACAAAA ATGAAAGTAG GGAAATGATT AACAATAATC ATAGGAATCT AACAATCCAA	2903
GTGGCTTGAG AATATTCATT CTTCTTGACA GTATAGATTC TTTACAATTT CGTAAGTTCC	2963
AATGTATGTT TTAGGAATAT GAGGTCATTA CTATTCATAA TCTGATACAG CTTTATCCTA	3023
AGGCCTCTCT TTAAAACTA CACTGCATCA TAGCTTTTTT GTGCAGTTGG TCTTTCTACT	3083
GTTACTGAAC AGTAAGCAAC CTACAGATTC ACTATCACCA ACCAGCCAGT TGATGGATCT	3143



TAAGCAAATT ATCAAGCTTG TGATAACCTA AATTATAAAA TGAGGGTGTG GGAATAGTTA	3203
CATTCCAAAT CTTCTATAAC ACTCTGTATT ATATTTCTGC CTCATTCCTT GTAG GGT Gly	3260
TTC TTC AGT GCC CGT GGT CAT CGA CCC CTT GAC AAG AAG AGA GAA GAG Phe Phe Ser Ala Arg Gly His Arg Pro Leu Asp Lys Lys Arg Glu Glu 40 45 50 55	3308
GCT CCC AGC CTG AGG CCT GCC CCA CCG CCC ATC AGT GGA GGT GGC TAT Ala Pro Ser Leu Arg Pro Ala Pro Pro Pro Ile Ser Gly Gly Gly Tyr 60 65 70	3356
CGG GCT CGT CCA GCC AAA GCA GCT GCC ACT CAA AAG AAA GTA GAA AGA Arg Ala Arg Pro Ala Lys Ala Ala Ala Thr Gln Lys Lys Val Glu Arg 75 80 85	3404
AAA GCC CCT GAT GCT GGA GGC TGT CTT CAC GCT GAC CCA GAC CTG Lys Ala Pro Asp Ala Gly Gly Cys Leu His Ala Asp Pro Asp Leu 90 95 100	3449
GTGGGTGCAC TGATGTTTCT TGCAGTGGTG GCTCTCTCAT GCAGAGAAAG CCTGTAGTCA	3509
TGGCAGTCTG CTAATGTTTC ACTGACCCAC ATTACCATCA CTGTTATTTT GTTTGTTTAT	3569
TTTGAAATA AAATTCAAAA CATAAACATA TTGGGCCTTT GGTTTAGGCT TTCTTTCTTG	3629
TTTTCTTTGG TCTGGGCCCA AAATTTCAAA TTAGGATATG TGGGTGCCAC CTTTCCATTT	3689
GTATTTTGCC ACTGCCTTTG TTTAGTTGGT AAAATTTTCA TAGCCCAATT ATATTTTTTC	3749
TGGGGTAAGT AATATTTTAA ATCTCTATGA GAGTATGATG ATGACTTTCG AATTTCTGGT	3809
CTTACAGAAA ACCAAATAAT AAATTTTTAT GTTGGCTAAT CGTATCGCTG AATTTTCCTA	3869
TGTGCTATTT TAACAAATGT CCATGACCCA AATCCTTCAT CTAATGCCTG CTATTTTCTT	3929
TGTTTTTAG GGG GTG TTG TGT CCT ACA GGA TGT CAG TTG CAA GAG GCT Gly Val Leu Cys Pro Thr Gly Cys Gln Leu Gln Glu Ala 105 110 115	3977
TTG CTA CAA CAG GAA AGG CCA ATC AGA AAT AGT GTT GAT GAG TTA AAT Leu Leu Gln Gln Glu Arg Pro Ile Arg Asn Ser Val Asp Glu Leu Asn 120 125 130	4025

AAC AAT GTG GAA GCT GTT TCC CAG ACC TCC TCT TCT TCC TTT CAG TAC	4073
Asn Asn Val Glu Ala Val Ser Gln Thr Ser Ser Ser Ser Phe Gln Tyr	
135 140 145	
ATG TAT TTG CTG AAA GAC CTG TGG CAA AAG AGG CAG AAG CAA GTA AAA G	4122
Met Tyr Leu Leu Lys Asp Leu Trp Gln Lys Arg Gln Lys Gln Val Lys	
150 155 160	
GTAGATATCC TTGTGCTTTC CATTGATTTC TCAGCTATAA AATTGGAACC GTTAGACTGC	4182
CACGAGAATG CATGGTTGTG AGAAGATTAA CATTCTCTGGG TTAGTGAATA GCATTCATAC	4242
GCTTTTGGGC ACCTTCCCCT GCAACTTGCC AGATAAGCAC TATTCAGCTC TTATTCCCAG	4302
TCTGACATCA GCAAGTGTGA TTTTCTATGA AAAATTCTAC TATGACTCCT TATTTTAAGT	4362
ATACAAGAAA CTTGTGACTC AGAAGATAAT ATTTACAGAG TGGAAAAAAA CCCCTAGCAT	4422
TTATAGTTTT AACATTTGAG GTTTTGAATG AGAGAGTTAT CCATAATATA TTCAATTGTG	4482
TTGTGGATAA TGACACCTAA CCTGTGAATC TTGAGGTCAG AATGTTGAGT GCTGTTGACT	4542
TGGTGGTCAG GAAACAGCTA GTGCGTGAGC CTGGCACAGG CATCTCAGTG AGTAGCATAC	4602
CCACAGTTGG AAATTTTTC AAGAAATCAA AGGAATCATG ACATCTTATA AATTTCAAGG	4662
TTCTGCTATA CTTATGTGAA ATGGATAAAT AAATCAAGCA TATCCACTCT GTAAGATTGA	4722
ACTTCTCAGA TGGAAGACCC CAATACTGCT TTCTCCTCTT TTCCCTCACC AAAGAAATAA	4782
ACAACCTATT TCATTTATTA CTGGACACAA TCTTTAGCGT ATACCTATGG TAAATTACTA	4842
GTATGGTGGT TAGGATTTAT GTTAATTTGT ATATGTCATG CGCCAAATCA TTTCCACTAA	4902
ATATGACTAT ATATCATAAC TGCTTGGTGA TAGCTCAGTG TTTAATAGTT TATTCTCAGA	4962
AAATCAAAAT TGTATAGTTA AATACATTAG TTTTATGAGG CAAAAATGCT AACTATTTCT	5022
ACATAATTTT ATTTTTCAG AT AAT GAA AAT GTA GTC AAT GAG TAC TCC	5071
Asp Asn Glu Asn Val Val Asn Glu Tyr Ser	
165 170	
TCA GAA CTG GAA AAG CAC CAA TTA TAT ATA GAT GAG ACT GTG AAT AGC	5119
Ser Glu Leu Glu Lys His Gln Leu Tyr Ile Asp Glu Thr Val Asn Ser	

175	180	185	
AAT ATC CCA ACT AAC CTT CGT GTG CTT CGT TCA ATC CTG GAA AAC CTG			5167
Asn Ile Pro Thr Asn Leu Arg Val Leu Arg Ser Ile Leu Glu Asn Leu			
190	195	200	205
AGA AGC AAA ATA CAA AAG TTA GAA TCT GAT GTC TCA GCT CAA ATG GAA			5215
Arg Ser Lys Ile Gln Lys Leu Glu Ser Asp Val Ser Ala Gln Met Glu			
	210	215	220
TAT TGT CGC ACC CCA TGC ACT GTC AGT TGC AAT ATT CCT GTG GTG TCT			5263
Tyr Cys Arg Thr Pro Cys Thr Val Ser Cys Asn Ile Pro Val Val Ser			
	225	230	235
GGC AAA G GTA ACTGATT CATAAACATA TTTT TAGAGA GTTCCAGAAG AACTCACACA			5320
Gly Lys			
CCAAAAATAA GAGAACAACA ACAACAACAA AAATGCTAAG TGGATTTTCC CAACAGATCA			5380
TAATGACATT ACAGTACATC ATAAAAATAT CCTTAGCCAG TTGTGTTTTG GACTGGCCTG			5440
GTGCATTTGC TGGTTTTGAT GAGCAGGATG GGGCACAGGT AGTCCCAGGG GTGGCTGATG			5500
TGTGCATCTG CGTACTGGCT TGAACAGATG GCAGAACCAC AGATAGATGT AGAAGTTTCT			5560
CCATTTTGTG TGTTCCTGGGA GCTCATGGAT ATTCCAGGAC ACAAAGGTG GAGAAGAGCT			5620
TTGTTTCATCC TCTTAGCAGA TAAACGTCCT CAAACTGGG TTGGACTTAC TAAAGTAAAA			5680
TGAAAATCTA ATATTTGTTA TATTATTTTC AAAGGTCTAT AATAACACAC TCCTTAGTAA			5740
CTTATGTAAT GTTATTTTAA AGAATTGGTG ACTAAATACA AAGTAATTAT GTCATAAACC			5800
CCTGAACATA ATGTTGTCTT ACATTTGCAG	AA TGT GAG GAA ATT ATC AGG AAA		5853
	Glu Cys Glu Glu Ile Ile Arg Lys		
	240	245	
GGA GGT GAA ACA TCT GAA ATG TAT CTC ATT CAA CCT GAC AGT TCT GTC			5901
Gly Gly Glu Thr Ser Glu Met Tyr Leu Ile Gln Pro Asp Ser Ser Val			
	250	255	260
AAA CCG TAT AGA GTA TAC TGT GAC ATG AAT ACA GAA AAT GGA G			5944
Lys Pro Tyr Arg Val Tyr Cys Asp Met Asn Thr Glu Asn Gly			
	265	270	275

GTAAGCTTTC GACAGTTGTT GACCTGTTGA TCTGTAATTA TTTGGATACC GTAAATGCC	6004
AGGAAACAAG GCCAGGTGTG GTGGCTCATA CCTGTAATTC CAGCACCTTG GGAGGCCAAA	6064
GTGGGCTGAT AGCTTGAGCC TAGGAGTTTG AACTAGCCT GGGCAACATA ATGAGACCCT	6124
AACTCTACAA AAAAAAAAAA AATACCAAAA AAAAAAAAAA AATCAGCTGT GTTGGTAGTA	6184
TGTGCCTGTA GTCCCAGCTA TCCAGGAGGC TGAGATGGGA GATCACCTGA GCCCACAACC	6244
TGGAGTCTTG ATCATGCTAC TGAAGTGTAG CCTGGGCAAC AGAGGATAGT GAGATCCTGT	6304
CTCAAAAAAAAA AAAATTAATT AAAAAGCCAG GAAACAAGAC TTAGCTCTAA CATCTAACAT	6364
AGCTGACAAA GGAGTAATTT GATGTGGAAT TCAACCTGAT ATTTAAAAGT TATAAATAT	6424
CTATAATTCA CAATTTGGGG TAAGATAAAG CACTTGCAGT TTCCAAAGAT TTTACAAGTT	6484
TACCTCTCAT ATTTATTTCC TTATTGTGTC TATTTTAGAG CACCAAATAT AACTAAATG	6544
GAATGGACAG GGGATTCAGA TATTATTTTC AAAGTGACAT TATTGCTGT TGGTTAATAT	6604
ATGCTCTTTT TGTTTCTGTC AACCAAAG GA TGG ACA GTG ATT CAG AAC CGT	6655
Gly Trp Thr Val Ile Gln Asn Arg	
280 285	
CAA GAC GGT AGT GTT GAC TTT GGC AGG AAA TGG GAT CCA TAT AAA CAG	6703
Gln Asp Gly Ser Val Asp Phe Gly Arg Lys Trp Asp Pro Tyr Lys Gln	
290 295 300	
GGA-TTT GGA AAT GTT GCA ACC AAC ACA GAT GGG AAG AAT TAC TGT GGC	6751
Gly Phe Gly Asn Val Ala Thr Asn Thr Asp Gly Lys Asn Tyr Cys Gly	
305 310 315	
CTA CCA G GTAACGAACA GGCATGCAAA ATAAAATCAT TCTATTTGAA ATGGGATTTT	6808
Leu Pro	
TTTTAATTAA AAAACATTCA TTGTTGGAAG CCTGTTTTAG GCAGTTAAGA GGAGTTTCCT	6868
GACAAAAATG TGGAAGCTAA AGATAAGGGA AGAAAGGCAG TTTTtagTTT CCCAAAATTT	6928
TATTTTGGT GAGAGATTTT ATTTTGTITT TCTTTTAG GT GAA TAT TGG CTT	6980
Gly Glu Tyr Trp Leu	

320

GGA AAT GAT AAA ATT AGC CAG CTT ACC AGG ATG GGA CCC ACA GAA CTT Gly Asn Asp Lys Ile Ser Gln Leu Thr Arg Met Gly Pro Thr Glu Leu 325 330 335 340	7028
TTG ATA GAA ATG GAG GAC TGG AAA GGA GAC AAA GTA AAG GCT CAC TAT Leu Ile Glu Met Glu Asp Trp Lys Gly Asp Lys Val Lys Ala His Tyr 345 350 355	7076
GGA GGA TTC ACT GTA CAG AAT GAA GCC AAC AAA TAC CAG ATC TCA GTG Gly Gly Phe Thr Val Gln Asn Glu Ala Asn Lys Tyr Gln Ile Ser Val 360 365 370	7124
AAC AAA TAC AGA GGA ACA GCC GGT AAT GCC CTC ATG GAT GGA GCA TCT Asn Lys Tyr Arg Gly Thr Ala Gly Asn Ala Leu Met Asp Gly Ala Ser 375 380 385	7172
CAG CTG ATG GGA GAA AAC AGG ACC ATG ACC ATT CAC AAC GGC ATG TTC Gln Leu Met Gly Glu Asn Arg Thr Met Thr Ile His Asn Gly Met Phe 390 395 400	7220
TTC AGC ACG TAT GAC AGA GAC AAT GAC GGC TG GTATGTGTGG Phe Ser Thr Tyr Asp Arg Asp Asn Asp Gly Trp 405 410 415	7262
CACTCTTTGC TCCTGCTTTA AAAATCACAC TAATATCATT ACTCAGAATC ATTAACAATA	7322
TTTTTAATAG CTACCACTTC CTGGGCACTT ACTGTCAGCC ACTGTCCTAA GCTCTTTATG	7382
CATCACTCGA AAGCATTTC ACTATAAGGT AGACATTCTT ATTCTCATTT TACAGATGAG	7442
ATTAGAGAG ATTACGTGAT TTGTCCAATG TCACACAAC ACCCAGAGAT AAAACTAGAA	7502
TTTGAGCACA GTTACTTTCT GAATAATGAG CATTAGATA AATACCTATA TCTCTATATT	7562
CTAAAGTGTG TGTGAAACT TTCATTTTCA TTTCCAGGGT TCTCTGATAC TAAGGGTTGT	7622
AAAAGCTATT ATTCCAGTAT AAAGTAACAA ACACAGTCCC TAGATGGATT GCCACAAAGG	7682
CCCAGTTATC TCTCTTTCTT GCTATAGGGC ACAGGAGGTC TTTGGTGTAT TAGTGTGACT	7742
CTATGTATAG CACCCAAAGG AAAGACTACT GTGCACACGA GTGTAGCAGT CTTTTATGGG	7802
TAATCTGCAA AACGTAAC TT GACCACCGTA GTTCTGTTTC TAATAACGCC AAACACATTT	7862

55

TCTTTCAG G TTA ACA TCA GAT CCC AGA AAA CAG TGT TCT AAA GAA GAC	7910
Leu Thr Ser Asp Pro Arg Lys Gln Cys Ser Lys Glu Asp	
420 425	
GGT GGT GGA TGG TGG TAT AAT AGA TGT CAT GCA GCC AAT CCA AAC GGC	7958
Gly Gly Gly Trp Trp Tyr Asn Arg Cys His Ala Ala Asn Pro Asn Gly	
430 435 440	
AGA TAC TAC TGG GGT GGA CAG TAC ACC TGG GAC ATG GCA AAG CAT GGC	8006
Arg Tyr Tyr Trp Gly Gly Gln Tyr Thr Trp Asp Met Ala Lys His Gly	
445 450 455 460	
ACA GAT GAT GGT GTA GTA TGG ATG AAT TGG AAG GGG TCA TGG TAC TCA	8054
Thr Asp Asp Gly Val Val Trp Met Asn Trp Lys Gly Ser Trp Tyr Ser	
465 470 475	
ATG AGG AAG ATG AGT ATG AAG ATC AGG CCC TTC TTC CCA CAG CAA TAGTCCCCAA	
8109	
Met Arg Lys Met Ser Met Lys Ile Arg Pro Phe Phe Pro Gln Gln	
480 485 490	
TACGTAGATT TTTGCTCTTC TGTATGTGAC AACATTTTTG TACATTATGT TATTGGAATT	8169
TTCTTTCATA CATTATATTC CTCTAAAACT CTCAAGCAGA CGTGAGTGTG ACTTTTTGAA	8229
AAAAGTATAG GATAAATTAC ATTAAAATAG CACATGATTT TCTTTTGTTT TCTTCATTTT	8289
TCTTGCTCAC CCAAGAAGTA ACAAAGTAT AGTTTTGACA GAGTTGGTGT TCATAATTTT	8349
AGTTCTAGTT GATTGCGAGA ATTTTCAAAT AAGGAAGAGG GGTCTTTTAT CCTTGTCGTA	8409
GGAAAACCAT GACGGAAAGG AAAAAGTAT GTTTAAAGT CCACTTTTAA AACTATATTT	8469
ATTTATGTAG GATCTGTCAA AGAAAAGTTC CAAAAGATT TATTAATTAA ACCAGACTCT	8529
GTTGCAATAA GTTAATGTTT TCTTGTTTTG TAATCCACAC ATTCAATGAG TTAGGCTTTG	8589
CACTTGTAAG GAAGGAGAAG CGTTCACAAC CTCAAATAGC TAATAAACCG GTCTTGAATA	8649
TTTGAAGATT TAAAATCTGA CTCTAGGACG GGCACGGTGG CTCACGACTA TAATCCCAAC	8709
ACTTTGGGAG GCTGAGGCGG GCGGTCACAA GGTCAGGAGT TCAAGACCAG CCTGACCAAT	8769
ATGGTGAAAC CCCATCTCTA CTAAAAATAC AAAAATTAGC CAGGCGTGGT GGCAGGTGCC	8829

TGTAGGTCCC AGCTAGCCTG TGAGGTGGAG ATTGCATTGA GCCAAGATC

8878

## (2) INFORMATION FOR SEQ ID NO:4:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 491 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Met	Lys	Arg	Met	Val	Ser	Trp	Ser	Phe	His	Lys	Leu	Lys	Thr	Met	Lys	1	5	10	15
His	Leu	Leu	Leu	Leu	Leu	Leu	Cys	Val	Phe	Leu	Val	Lys	Ser	Gln	Gly	20	25	30	
Val	Asn	Asp	Asn	Glu	Glu	Gly	Phe	Phe	Ser	Ala	Arg	Gly	His	Arg	Pro	35	40	45	
Leu	Asp	Lys	Lys	Arg	Glu	Glu	Ala	Pro	Ser	Leu	Arg	Pro	Ala	Pro	Pro	50	55	60	
Pro	Ile	Ser	Gly	Gly	Gly	Tyr	Arg	Ala	Arg	Pro	Ala	Lys	Ala	Ala	Ala	65	70	75	80
Thr	Gln	Lys	Lys	Val	Glu	Arg	Lys	Ala	Pro	Asp	Ala	Gly	Gly	Cys	Leu	85	90	95	
His	Ala	Asp	Pro	Asp	Leu	Gly	Val	Leu	Cys	Pro	Thr	Gly	Cys	Gln	Leu	100	105	110	
Gln	Glu	Ala	Leu	Leu	Gln	Gln	Glu	Arg	Pro	Ile	Arg	Asn	Ser	Val	Asp	115	120	125	
Glu	Leu	Asn	Asn	Asn	Val	Glu	Ala	Val	Ser	Gln	Thr	Ser	Ser	Ser	Ser	130	135	140	
Phe	Gln	Tyr	Met	Tyr	Leu	Leu	Lys	Asp	Leu	Trp	Gln	Lys	Arg	Gln	Lys	145	150	155	160

Gln Val Lys Asp Asn Glu Asn Val Val Asn Glu Tyr Ser Ser Glu Leu  
 165 170 175  
 Glu Lys His Gln Leu Tyr Ile Asp Glu Thr Val Asn Ser Asn Ile Pro  
 180 185 190  
 Thr Asn Leu Arg Val Leu Arg Ser Ile Leu Glu Asn Leu Arg Ser Lys  
 195 200 205  
 Ile Gln Lys Leu Glu Ser Asp Val Ser Ala Gln Met Glu Tyr Cys Arg  
 210 215 220  
 Thr Pro Cys Thr Val Ser Cys Asn Ile Pro Val Val Ser Gly Lys Glu  
 225 230 235 240  
 Cys Glu Glu Ile Ile Arg Lys Gly Gly Glu Thr Ser Glu Met Tyr Leu  
 245 250 255  
 Ile Gln Pro Asp Ser Ser Val Lys Pro Tyr Arg Val Tyr Cys Asp Met  
 260 265 270  
 Asn Thr Glu Asn Gly Gly Trp Thr Val Ile Gln Asn Arg Gln Asp Gly  
 275 280 285  
 Ser Val Asp Phe Gly Arg Lys Trp Asp Pro Tyr Lys Gln Gly Phe Gly  
 290 295 300  
 Asn Val Ala Thr Asn Thr Asp Gly Lys Asn Tyr Cys Gly Leu Pro Gly  
 305 310 315 320  
 Glu Tyr Trp Leu Gly Asn Asp Lys Ile Ser Gln Leu Thr Arg Met Gly  
 325 330 335  
 Pro Thr Glu Leu Leu Ile Glu Met Glu Asp Trp Lys Gly Asp Lys Val  
 340 345 350  
 Lys Ala His Tyr Gly Gly Phe Thr Val Gln Asn Glu Ala Asn Lys Tyr  
 355 360 365  
 Gln Ile Ser Val Asn Lys Tyr Arg Gly Thr Ala Gly Asn Ala Leu Met  
 370 375 380  
 Asp Gly Ala Ser Gln Leu Met Gly Glu Asn Arg Thr Met Thr Ile His  
 385 390 395 400



Asn	Gly	Met	Phe	Phe	Ser	Thr	Tyr	Asp	Arg	Asp	Asn	Asp	Gly	Trp	Leu
			405						410					415	
Thr	Ser	Asp	Pro	Arg	Lys	Gln	Cys	Ser	Lys	Glu	Asp	Gly	Gly	Gly	Trp
			420					425					430		
Trp	Tyr	Asn	Arg	Cys	His	Ala	Ala	Asn	Pro	Asn	Gly	Arg	Tyr	Tyr	Trp
		435					440					445			
Gly	Gly	Gln	Tyr	Thr	Trp	Asp	Met	Ala	Lys	His	Gly	Thr	Asp	Asp	Gly
	450					455					460				
Val	Val	Trp	Met	Asn	Trp	Lys	Gly	Ser	Trp	Tyr	Ser	Met	Arg	Lys	Met
465					470					475					480
Ser	Met	Lys	Ile	Arg	Pro	Phe	Phe	Pro	Gln	Gln					
			485						490						

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10564 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA (genomic)

(vii) IMMEDIATE SOURCE:

- (B) CLONE: human fibrinogen gamma chain

(ix) FEATURE:

- ```
(A) NAME/KEY: CDS
(B) LOCATION: join(1799..1876, 1973..2017, 2207..2390, 2510
    ..2603, 4211..4341, 4645..4778, 5758..5942, 7426
    ..7703, 9342..9571)
```

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

CTACACACTT CTTGAAGGCA AAGGCAATGC TGAAGTCACC TTTCATGTTC AAATCATATT 60

AAAAAGTTAG CAAGATGTAA TTATCAGTGT ACTATGTAAA TCTTTGTGAA TGATCAATAA 120

|                                                                   |      |
|-------------------------------------------------------------------|------|
| TTACATATTT TCATTATATA TATTTTAGTA GATAATATTT ATATACATTC AACATTCTAA | 180  |
| ATATAGAAAG TTTACAGAGA AAAATAAAGC CTTTTTTTCC AATCCTGTCC TCCACCTCTG | 240  |
| CATCCCATTC TTCTTCACAG AGGCAACTGA TTCAAGTCAT TACATAGTTA TTGAGTGTTA | 300  |
| ACTACAATA TGTAAAGTAC AGCTATATAT GTTAGATGCC GTAGCCACAG AAATCAGTTT  | 360  |
| ACAATCTAAT GCAGTGGATA CAGCATGTAT ACATATAATA TAAGGTTGCT ACAAATGCTA | 420  |
| TCTGAGGTAG AGCTGTTTGA AAGAATACTA ATACTTAAAT GTTTAATTCA ACTGACTTGA | 480  |
| TTGACAAC TGATTAGCTGAG TGGAAAAGAT GGATGAGAAA GATTGTGAGA CTTAATTGGC | 540  |
| TGGTGGTATG GTGATATGAT TGACAATAAC TGCTAAGTCA GAGAGGGATA TATTAAGGAG | 600  |
| GAGAAGAAAA GCAACAAATC TGGTTTTGAT GTGTTCACTT TGTATAATT ATTGATTATT  | 660  |
| TACTGAATAT GAATATTTAT CTTTGTTTTT GAGTCAATAA ATATACCTTT GTAAAGACAG | 720  |
| AATTAAAGTA TTAGTATTTT TTTCAAAC TGAGGCATTT TCCCACTAAC ATATTTTCATC  | 780  |
| AAAACCTATA ATAAGCTTGG TTCCAGAGGA AGAAATGAGG GATAACCAAA AATAGAGACA | 840  |
| TTAATAATAG TGTAACGCCC AGTGATAAAT CTCAATAGGC AGTGATGACA GACATGTTTT | 900  |
| CCCCAACACA AGGATGCTGT AAGGGCCAAA CAGAAATGAT GGCCCTCCC CAGCACCTCA  | 960  |
| TTTTGCCCCT TCCTTCAGCT ATGCCTCTAC TCTCCTTTAG ATACAAGGGA GGTGGATTTT | 1020 |
| TCTCTTCTCT GAGATAGCTT GATGGAACCA CAGGAACAAT GAAGTGGGCT CCTGGCTCTT | 1080 |
| TTCTCTGTGG CAGATGGGGT GCCATGCCCA CCTTCAGACA AAGGGAAGAT TGAGCTCAAA | 1140 |
| AGCTCCCTGA GAAGTGAGAG CCTATGAACA TGGTTGACAC AGAGGGACAG GAATGTATTT | 1200 |
| CCAGGGTCAT TCATTCCTGG GAATAGTGAA CTGGGACATG GGGGAAGTCA GTCTCCTCCT | 1260 |
| GCCACAGCCA CAGATTAAAA ATAATAATGT TAACTGATCC CTAGGCTAAA ATAATAGTGT | 1320 |
| TAACTGATCC CTAAGCTAAG AAAGTTCTTT TGGTAATTCA GGTGATGGCA GCAGGACCCA | 1380 |
| TCTTAAGGAT AGACTAGGTT TGCTTAGTTC GAGGTCATAT CTGTTTGCTC TCAGCCATGT | 1440 |

ACTGGAAGAA GTTGCATCAC ACAGCCTCCA GGA CTGCCCT CCTCCTCACA GCAATGGATA 1500  
 ATGCTTCACT AGCCTTTGCA GATAATTTTG GATCAGAGAA AAAACCTTGA GCTGGGCCAA 1560  
 AAAGGAGGAG CTTCAACCTG TGTGCAAAAT CTGGGAACCT GACAGTATAG GTTGGGGGCC 1620  
 AGGATGAGGA AAAAGGAACG GGAAAGACCT GCCCACCCTT CTGGTAAGGA GGCCCCGTGA 1680  
 TCAGCTCCAG CCATTTGCAG TCCTGGCTAT CCCAGGAGCT TACATAAAGG GACAATTGGA 1740  
 GCCTGAGAGG TGACAGTGCT GACACTACAA GGCTCGGAGC TCCGGGCACT CAGACATC 1798  
 ATG AGT TGG TCC TTG CAC CCC CGG AAT TTA ATT CTC TAC TTC TAT GCT 1846  
 Met Ser Trp Ser Leu His Pro Arg Asn Leu Ile Leu Tyr Phe Tyr Ala  
 1 5 10 15  
 CTT TTA TTT CTC TCT TCA ACA TGT GTA GCA GTAAGTGTGC TCTTCACAAA 1896  
 Leu Leu Phe Leu Ser Ser Thr Cys Val Ala  
 20 25  
 ACGTTGTTTA AAATGGAAAG CTGGAAAATA AAACAGATAA TAAACTAGTG AAATTTTCGT 1956  
 ATTTTTTCTC TTTTAG TAT GTT GCT ACC AGA GAC AAC TGC TGC ATC TTA 2005  
 Tyr Val Ala Thr Arg Asp Asn Cys Cys Ile Leu  
 30 35  
 GAT GAA AGA TTC GTAAGTAGTT TTTATGTTTC TCCCTTTGTG TGTGAACTGG 2057  
 Asp Glu Arg Phe  
 40  
 AGAGGGGCAG AGGAATAGAA ATAATTCCT CATAAATATC ATCTGGCACT TGTAACTTTT 2117  
 TAAAAACATA GTCTAGGTTT TACCTATTTT TCTTAATAGA TTTTAAGAGT AGCATCTGTC 2177  
 TACATTTTTA ATCACTGTTA TATTTTCAG GGT AGT TAT TGT CCA ACT ACC TGT 2230  
 Gly Ser Tyr Cys Pro Thr Thr Cys  
 45  
 GGC ATT GCA GAT TTC CTG TCT ACT TAT CAA ACC AAA GTA GAC AAG GAT 2278  
 Gly Ile Ala Asp Phe Leu Ser Thr Tyr Gln Thr Lys Val Asp Lys Asp  
 50 55 60 65  
 CTA CAG TCT TTG GAA GAC ATC TTA CAT CAA GTT GAA AAC AAA ACA TCA 2326  
 Leu Gln Ser Leu Glu Asp Ile Leu His Gln Val Glu Asn Lys Thr Ser  
 70 75 80

|                                                                                                                                                   |      |
|---------------------------------------------------------------------------------------------------------------------------------------------------|------|
| GAA GTC AAA CAG CTG ATA AAA GCA ATC CAA CTC ACT TAT AAT CCT GAT<br>Glu Val Lys Gln Leu Ile Lys Ala Ile Gln Leu Thr Tyr Asn Pro Asp<br>85 90 95    | 2374 |
| GAA TCA TCA AAA CCA A GTGAGAAAAT AAAGACTACT GACCAAAAAA<br>Glu Ser Ser Lys Pro<br>100                                                              | 2420 |
| TAATAATAAT AATCTGTGAA GTTCTTTTGC TGTGTGTTTAA GTTGTCTCTAT TTGCTTAAGG                                                                               | 2480 |
| ATTTTTATGT CTCTGATCCT ATATTACAG AT ATG ATA GAC GCT GCT ACT TTG<br>Asn Met Ile Asp Ala Ala Thr Leu<br>105 110                                      | 2532 |
| AAG TCC AGG ATA ATG TTA GAA GAA ATT ATG AAA TAT GAA GCA TCG ATT<br>Lys Ser Arg Ile Met Leu Glu Glu Ile Met Lys Tyr Glu Ala Ser Ile<br>115 120 125 | 2580 |
| TTA ACA CAT GAC TCA AGT ATT CG GTAAGGATTT TTGTTTTAAT TTGCTCTGCA<br>Leu Thr His Asp Ser Ser Ile Arg<br>130                                         | 2633 |
| AGACTGATTT AGTTTTTATT TAATATTCTA TACTTGAGTG AAAGTAATTT TTAATGTGTT                                                                                 | 2693 |
| TTCCCCATTT ATAATATCCC AGTGACATTA TGCCTGATTA TGTGAGCAT AGTAGAGATA                                                                                  | 2753 |
| GAAGTTTTTA GTGCAATATA AATTATACTG GGTATAAATT GCTTATTAAT AATCACATTG                                                                                 | 2813 |
| AAGAAAGATG TTCTAGATGT CTTCAAATGC TAGTTTGACC ATATTTATCA AAAATTTTTT                                                                                 | 2873 |
| CCCCATCCCC CATTTATCTT ACAACATAAA ATCAATCTCA TAGGAATTTG GGTGTTGAAA                                                                                 | 2933 |
| ATAAAATCCT CTTTATAAAA ATGCTGACAA ATTGGTGGTT AAAAAAATTA GCAAGCAGAG                                                                                 | 2993 |
| GCATAGTAAG GATTTTGGCT CCTAAAGTAA ATTATATTGA ATGTGGAGCA GGAAGAAACA                                                                                 | 3053 |
| TGTCTTGAGA GACTAAGTGT GGCAAATATT GCAAAGCTCA TATTGATCAT TGCAGAATGA                                                                                 | 3113 |
| ACCTGCATAG TCTCTTCCT TCATTTGGAA GTGAATGTCT CTGTAAAGC TTCTCAGGGA                                                                                   | 3173 |
| CTCATAACT TTCTGAACAT AAGGTCTCAG ATACAGTTTT AATATTTTTC CCAATTTTTT                                                                                  | 3233 |
| TTTTCTGAAT TTTTCTCAA GCAGCTTGAG AAATTGAGAT AAATAGTAGC TAGGGAGAAG                                                                                  | 3293 |

|                                                                    |      |
|--------------------------------------------------------------------|------|
| TGGCCCAGGA AAGATTTCTC CTCTTTTTCG TATCAGAGGG CCCTTGTTAT TATTGTTATT  | 3353 |
| ATTATTACTT GCATTATTAT TGTCCATCAT TGAAGTTGAA GGAGGTTATT GTACAGAAAT  | 3413 |
| TGCCTAAGAC AAGGTAGAGG GAAAACGTGG ACAAATAGTT TGTCTACCCT TTTTACTTC   | 3473 |
| AAAGAAAGAA CGGTTTATGC ATTGTAGACA GTTTTCTATC ATTTTGGAT ATTTGCAAGC   | 3533 |
| CACCCTGTAA GTAAC TACAA AAGGAGGGTT TTTACTTCCC CCAGTCCATT CCCAAAGCTA | 3593 |
| TGTAACCAGA AGCATTAAAG AAGAAAGGGG AAGTATCTGT TGTTTTATTT TACATACAAT  | 3653 |
| AACGTTCCAG ATCATGTCCC TGTGTAAGTT ATATTTTAGA TTGAAGCTTA TATGTATAGC  | 3713 |
| CTCAGTAGAT CCACAAGTGA AAGGTATACT CCTTCAGCAC ATGTGAATTA CTGAACTGAG  | 3773 |
| CTTTTCCTGC TTCTAAAGCA TCAGGGGGTG TTCCTATTAA CCAGTCTCGC CACTCTTGCA  | 3833 |
| GGTTGCTATC TGCTGTCCCT TATGCATAAA GTAAAAAGCA AAATGTCAAT GACATTTGCT  | 3893 |
| TATTGACAAG GACTTTGTTA TTTGTGTTGG GAGTTGAGAC AATATGCCCC ATTCTAAGTA  | 3953 |
| AAAAGATTCA GGTCCACATT GTATTCCTGT TTTAATTGAT TTTTGGATTT GTTTTCTTT   | 4013 |
| TTCAAAAAGT TTATAATTTT AATTCATGTT AATTTAGTAA TATAATTTA CATTTTCCTC   | 4073 |
| AAGAATGGAA TAATTTATCA GAAAGCACTT CTTAAGAAAA TACTTAGCAG TTTCCAAAGA  | 4133 |
| AAATATAAAA TTA CTCTTCT GAAAGGAATA CTTATTTTGG TCTTCTTATT TTTGTTATCT | 4193 |
| TATGTTTCTG TTTGTAG A TAT TTG CAG GAA ATA TAT AAT TCA AAT AAT CAA   | 4244 |
| Tyr Leu Gln-Glu Ile Tyr Asn Ser Asn Asn Gln                        |      |
| 135 140 145                                                        |      |
| AAG ATT GTT AAC CTG AAA GAG AAG GTA GCC CAG CTT GAA GCA CAG TGC    | 4292 |
| Lys Ile Val Asn Leu Lys Glu Lys Val Ala Gln Leu Glu Ala Gln Cys    |      |
| 150 155 160                                                        |      |
| CAG GAA CCT TGC AAA GAC ACG GTG CAA ATC CAT GAT ATC ACT GGG AAA G  | 4341 |
| Gln Glu Pro Cys Lys Asp Thr Val Gln Ile His Asp Ile Thr Gly Lys    |      |
| 165 170 175                                                        |      |
| GTAAC TGATG AAGGTTATAT TGGGATTAGG TTCATCAAAG TAAGTAATGT AAAGGAGAAA | 4401 |
| GTATGTACTG GAAAGTATAG GAATAGTTTA GAAAGTGGCT ACCCATTAAG TCTAAGAATT  | 4461 |

|                                                                    |      |
|--------------------------------------------------------------------|------|
| TCAGTTGTCT AGACCTTTCT TGAATAGCTA AAAAAACAG TTTAAAAGGA ATGCTGATGT   | 4521 |
| GAAAAGTAAG AAAATTATTC TTGGAAAATG AATAGTTTAC TACATGTTAA AAGCTATTTT  | 4581 |
| TCAAGGCTGG CACAGTCTTA CCTGCATTTT AAACCACAGT AAAAGTCGAT TCTCCTTCTC  | 4641 |
| TAG AT TGT CAA GAC ATT GCC AAT AAG GGA GCT AAA CAG AGC GGG CTT     | 4688 |
| Asp Cys Gln Asp Ile Ala Asn Lys Gly Ala Lys Gln Ser Gly Leu        |      |
| 180 185 190                                                        |      |
| TAC TTT ATT AAA CCT CTG AAA GCT AAC CAG CAA TTC TTA GTC TAC TGT    | 4736 |
| Tyr Phe Ile Lys Pro Leu Lys Ala Asn Gln Gln Phe Leu Val Tyr Cys    |      |
| 195 200 205                                                        |      |
| GAA ATC GAT GGG TCT GGA AAT GGA TGG ACT GTG TTT CAG AAG            | 4778 |
| Glu Ile Asp Gly Ser Gly Asn Gly Trp Thr Val Phe Gln Lys            |      |
| 210 215 220                                                        |      |
| GTAATTTTTT CCCCACCATG TGTATTTAAT AAATTCCTAC ATTGTTTCTG CCATATGGCA  | 4838 |
| GATACTTTTC TAAGCACCTT GTGAACCGTA GCTCATTTAA TCCTTGCAAT AGCCCTAAGA  | 4898 |
| GGAAGGTACT TCTGTTACTC CTATTTACAG AAAAGGAAAC TGAGGCACAC AAGGTTAAAT  | 4958 |
| AACTTGCCCA AGACCACATA ACTAATAAGC AACAGAGTCA GCATTTGAAC CTAGGCAGTA  | 5018 |
| TAGTTTCAGA GTTTGTGACT TGA CTCTATA TTGTACTGGC ACTGACTTTG TAGATTCATG | 5078 |
| GTGGCACATA ATCATAGTAC CACAGTGACA AATAAAAAGA AGGAAACTCT TTTGTCAGGT  | 5138 |
| AGGTCAAGAC CTGAGGTTTC CCATCACAAG ATGAGGAAGC CCAACACCAC CCCCACCAC   | 5198 |
| CCCACCACCA TCACCACCCT TTCACACACC AGAGGATACA CTTGGGCTGC TCCAAGACAA  | 5258 |
| GGAACCTGTG TTGCATCTGC CACTTGCTGA TACCCACTAG GAATCTTGGC TCCTTTACTT  | 5318 |
| TCTGTTTACC TCCCACCACT GTTATAACTG TTTCTACAGG GGGCGCTCAG AGGGAATGAA  | 5378 |
| TGGTGGAAGC ATTAGTTGCC AGACACCGAT TGAGCAATGG GTTCCATCAT AAGTGTAAGA  | 5438 |
| ATCAGTAATA TCCAGCTAGA GTTCTGAAGT CGTCTAGGTG TCTTTTAAAT ATTACCACTC  | 5498 |
| ATTTAGAATT TATGATGTGC CAGAAACCCT CTTAAGTATT TCTCTTATAT TCTCTCTCAT  | 5558 |

|                                                                    |      |
|--------------------------------------------------------------------|------|
| GATCCTTGCA GCAACCCTAA GAAGTAACCA TCATTTTTCC TATTTGATAC ATGAGGAAAC  | 5618 |
| TGAGGTAGCT TGGCCAAGAT CACTTAGTTG GGAGTTGATA GAACCAGTGC TCTGTATTTT  | 5678 |
| TGACAAAATG TTGACAGCAT TCTCTTTACA TGCATTGATA GTCTATTTTC TCCTTTTGCT  | 5738 |
| CTTGCAAATG TGTAATTAG AGA CTT GAT GGC AGT GTA GAT TTC AAG AAA AAC   | 5790 |
| Arg Leu Asp Gly Ser Val Asp Phe Lys Lys Asn                        |      |
| 225 230                                                            |      |
| TGG ATT CAA TAT AAA GAA GGA TTT GGA CAT CTG TCT CCT ACT GGC ACA    | 5838 |
| Trp Ile Gln Tyr Lys Glu Gly Phe Gly His Leu Ser Pro Thr Gly Thr    |      |
| 235 240 245                                                        |      |
| ACA GAA TTT TGG CTG GGA AAT GAG AAG ATT CAT TTG ATA AGC ACA CAG    | 5886 |
| Thr Glu Phe Trp Leu Gly Asn Glu Lys Ile His Leu Ile Ser Thr Gln    |      |
| 250 255 260 265                                                    |      |
| TCT GCC ATC CCA TAT GCA TTA AGA GTG GAA CTG GAA GAC TGG AAT GGC    | 5934 |
| Ser Ala Ile Pro Tyr Ala Leu Arg Val Glu Leu Glu Asp Trp Asn Gly    |      |
| 270 275 280                                                        |      |
| AGA ACC AG GTACTGTTTT GAAATGACTT CCAACTTTTT ATTGTAAAGA             | 5982 |
| Arg Thr Ser                                                        |      |
| TTGCCTGGAA TGTGCACTTT CCAACTATCA ATAGACAATG GCAAATGCAG CCTGACAAAT  | 6042 |
| GCAAACAGCA CATCCAGCCA CCATTTTCTC CAGGAGTCTG TTTGGTTCTT GGGCAATCCA  | 6102 |
| AAAAGGTAAA TTCTATTGAG GATGAATCTA AGTGTATTGG TACAATCTAA TTACCCTGGA  | 6162 |
| ACCATTGAGA GTAATAGCTA ATTACTGAAC TTTTAATCAG TCCCAGGAAT TGAGCATAAA  | 6222 |
| ATTATAATTT TATCTAGTCT AAATTACTAT TTCATGAAGC AGGTATTATT ATTAATCCCA  | 6282 |
| TTTTATAGAT TAACTTGCTC AAAGTCACAT TGCTGATAAG TGGTAGAGGT AGAATTCAGA  | 6342 |
| CTCAAGTAGT TTAACCTTAG AGCCTGTCCT CTTAACAACCT ATCCTGGTTG AAAAGCAAAT | 6402 |
| ACAGCCTCTT CAGACTTCTC AGTGCCTTGA TGGCCATTTA TTCTGTCAAA TCATGAGCTA  | 6462 |
| CCCTAAAAGT AAACCAGCTA GCTCTTTTGA TGATCTAGAG GCTTCTTTTT GCTTGAGATA  | 6522 |
| TTTGAAGGTT TTAAGCATTG TTACCTAATT AAAATGCAGA AAAATATCCA ACCCTCTTGT  | 6582 |

65

|                                                                   |      |
|-------------------------------------------------------------------|------|
| TATGTTTAAG GAATAGTGAA ATATATTGTC TTCAAACACA TGGACTTTTT TTTATTGCTT | 6642 |
| GGTTGGTTTT TAATCCAGAA AGTGCTATAG TCAGTAGACC TTCTTCTAGG AAAGGACCTT | 6702 |
| CCATTTCCCA GCCACTGGAG ATTAGAAAAT AAGCTAAATA TTTTCTGGAA ATTTCTGTTC | 6762 |
| ATTCATTAAG GCCCATCCTT TCCCCCACTC TATAGAAGTG TTGTCCACTT GCACAATTTT | 6822 |
| TTCCAGGAAA GAATCTCTCT AACTCCTTCA GCTCACATGC TTTGGACCAC ACAGGGAAGA | 6882 |
| CTTTGATTGT GTAATGCCCT CAGAAGCTCT CCTTCTTGCC ACTACCACAC TGATTTGAGG | 6942 |
| AAGAAAATCC CTTTAGCACC TAACCCTTCA GGTGCTATGA GTGGCTAATG GAACTGTACC | 7002 |
| TCCTTCAAGT TTTGTGCAAT AATTAAGGGT CACTCACTGT CAGATACTTT CTGTGATCTA | 7062 |
| TGATAATGTG TGTGCAACAC ATAACATTTT AATAAAAGTA GAAAATATGA AATTAGAGTC | 7122 |
| ATCTACACAT CTGGATTTGA TCTTAGAATG AAACAAGCAA AAAAGCATCC AAGTGAGTGC | 7182 |
| AATTATTAGT TTTCAGAGAT GCTTCAAAGG CTTCTAGGCC CATCCCGGGA AGTGTTAATG | 7242 |
| AGCTGTGGAC TGGTTCACAT ATCTATTGCC TCTTGCCAGA TTTGCAAAAA ACTTCACTCA | 7302 |
| ATGAGCAAAT TTCAGCCTTA AGAAACAAAG TCAAAAATTC CAAGGAAGCA TCCTACGAAA | 7362 |
| GAGGGAACCT CTGAGATCCC TGAGGAGGGT CAGCATGTGA TGGTTGTATT TCCTTCTTCT | 7422 |
| CAG T ACT GCA GAC TAT GCC ATG TTC AAG GTG GGA CCT GAA GCT GAC     | 7468 |
| Thr Ala Asp Tyr Ala Met Phe Lys Val Gly Pro Glu Ala Asp           |      |
| 285 290 295                                                       |      |
| AAG TAC CGC CTA ACA TAT GCC TAC TTC GCT GGT GGG GAT GCT GGA GAT   | 7516 |
| Lys Tyr Arg Leu Thr Tyr Ala Tyr Phe Ala Gly Gly Asp Ala Gly Asp   |      |
| 300 305 310                                                       |      |
| GCC TTT GAT GGC TTT GAT TTT GGC GAT GAT CCT AGT GAC AAG TTT TTC   | 7564 |
| Ala Phe Asp Gly Phe Asp Phe Gly Asp Asp Pro Ser Asp Lys Phe Phe   |      |
| 315 320 325 330                                                   |      |
| ACA TCC CAT AAT GGC ATG CAG TTC AGT ACC TGG GAC AAT GAC AAT GAT   | 7612 |
| Thr Ser His Asn Gly Met Gln Phe Ser Thr Trp Asp Asn Asp Asn Asp   |      |
| 335 340 345                                                       |      |



|                                                                    |      |
|--------------------------------------------------------------------|------|
| AAG TTT GAA GGC AAC TGT GCT GAA CAG GAT GGA TCT GGT TGG TGG ATG    | 7660 |
| Lys Phe Glu Gly Asn Cys Ala Glu Gln Asp Gly Ser Gly Trp Trp Met    |      |
| 350 355 360                                                        |      |
| AAC AAG TGT CAC GCT GGC CAT CTC AAT GGA GTT TAT TAC CAA G          | 7703 |
| Asn Lys Cys His Ala Gly His Leu Asn Gly Val Tyr Tyr Gln            |      |
| 365 370 375                                                        |      |
| GTATGTTTTTC CTTTCTTAGA TTCCAAGTTA ATGTATAGTG TATACTATTT TCATAAAAAA | 7763 |
| TAATAAATAG ATATGAAGAA ATGAAGAATA ATTTATAAAG ATAGTAGGGA TTTTATCATG  | 7823 |
| TTCTTTATTT CAACTAAGTT CTTTGAAACT GGAAGTGGAT AATACCAAGT TCATGCCTAA  | 7883 |
| AATTAGCCCT TCTAAAGAAA TCCACCTGCT GCAAATATC CAGTAGTTTG GCATTATATG   | 7943 |
| TGAAACTATC ACCATCATAG CTGGCACTGT GGGTTGTGGG ATCTCCTTTA GACATACAAC  | 8003 |
| ATAAATGATC TGGATGGATT AACATTACTA CATGGATGCT TGTGACACA TTAACCTGGC   | 8063 |
| TTCCCATGAG CTTTGTGTCA GATACACGCA GTGAACAGGT GTTTGGAGGA ACAGAATAAA  | 8123 |
| GAGAAGGCAA GCACTGGTAA GGGCAGGGGT TTGTGAAAGC TTGAGAGAAG AGACCAGTCT  | 8183 |
| GAGGACAGTA GACACTTATT TTAGGATGGG GGTTGGATGA GGAGGCTATA GTTTGCTATA  | 8243 |
| AGCTTGGAAT GGTTTGGAAC ACTGGTTTCA CTCACCTACC CAGCAGTTAT GTGTGGGGAA  | 8303 |
| GCCTTACCGA TGCTAAAGGA TCCATGTTAC AATAATGGCA TTATTTGGAA ATCCCAGTGG  | 8363 |
| TATTCCATGA ATAAAACCAC TATGAAGATA ATCCCACTCA ACAGACTCTC CGTTGGAGAA  | 8423 |
| GGACAGCAAC ACCACCCTGG GAAAGCCAAA CAGTCAGACC AGACCTGTTT AGCATCAGTA  | 8483 |
| GGACTTCCCT ACCATATCTG CTGGGTAGAT GAGTGAAACC AGTGTTCCAA ACCACTCCGG  | 8543 |
| GCTTGTAGCA AACCATAGTC TCCTCATCTA CCAAGATGAG CAACCTTACC TCCTGATGTC  | 8603 |
| CTAGCCAATC ACCAACTAGG AAACTTTGCA CAGTTTATTT AAAGTAACAG TTTGATTTTC  | 8663 |
| ACAATATTTT TAAATTGGAG AAACATAACT TATCTTTGCA CTCACAAACC ACATAATGAG  | 8723 |
| AAGAACTCT AAGGGAAAAT GCTTGATCTG TGTGACCCGG GGCGCCATGC CAGAGCTGTA   | 8783 |
| GTTTCATGCCA GTGTTGTGCT CTGACAAGCC TTTTACAGAA TTACATGAGA TCTGCTTCCC | 8843 |

|                                                                    |      |
|--------------------------------------------------------------------|------|
| TAGGACAAGG AGAAGGCAAA TCAACAGAGG CTGCACTTTA AAATGGAGAC ATAAAATAAC  | 8903 |
| ATGCCAGAAC CATTTCTCTAA AGCTCCTCAA TCAACCAACA AAATTGTGCT TTCAAATAAC | 8963 |
| CTGAGTTGAC CTCATCAGGA ATTTTGTGGC TCCTTCTCTT CTAACCTGCC TGAAGAAAGA  | 9023 |
| TGGTCCACAG CAGCTGAGTC CGGGATGGAT AAGCTTAGGG ACAGAGGCCA ATTAGGGAAC  | 9083 |
| TTTGGGTTTC TAGCCCTACT AGTAGTGAAT AAATTTAAAG TGTGGATGTG ACTATGAGTC  | 9143 |
| ACAGCACAGA TGTTGTTTAA TAATATGTTT ATTTTATAAA TTGATATTTT AGGAATCTTT  | 9203 |
| GGAGATATTT TCAGTTAGCA GATAATACTA TAAATTTTAT GTAACCTGGCA ATGCACTTCG | 9263 |
| TAATAGACAG CTCTTCATAG ACTTGCAGAG GTAAAAAGAT TCCAGAATAA TGATATGTAC  | 9323 |
| ATCTACGACT TGTTTTAG GT GGC ACT TAC TCA AAA GCA TCT ACT CCT AAT     | 9373 |
| Gly Gly Thr Tyr Ser Lys Ala Ser Thr Pro Asn                        |      |
| 380 385                                                            |      |
| GGT TAT GAT AAT GGC ATT ATT TGG GCC ACT TGG AAA ACC CGG TGG TAT    | 9421 |
| Gly Tyr Asp Asn Gly Ile Ile Trp Ala Thr Trp Lys Thr Arg Trp Tyr    |      |
| 390 395 400                                                        |      |
| TCC ATG AAG AAA ACC ACT ATG AAG ATA ATC CCA TTC AAC AGA CTC ACA    | 9469 |
| Ser Met Lys Lys Thr Thr Met Lys Ile Ile Pro Phe Asn Arg Leu Thr    |      |
| 405 410 415                                                        |      |
| ATT GGA GAA GGA CAG CAA CAC CAC CTG GGG GGA GCC AAA CAG GTC AGA    | 9517 |
| Ile Gly Glu Gly Gln Gln His His Leu Gly Gly Ala Lys Gln Val Arg    |      |
| 420 425 430 435                                                    |      |
| CCA GAG CAC CCT GCG GAA ACA GAA TAT GAC TCA CTT TAC CCT GAG GAT    | 9565 |
| Pro Glu His Pro Ala Glu Thr Glu Tyr Asp Ser Leu Tyr Pro Glu Asp    |      |
| 440 445 450                                                        |      |
| GAT TTG TAGAAAATTA ACTGCTAACT TCTATTGACC CACAAAGTTT CAGAAATTCT     | 9621 |
| Asp Leu                                                            |      |
| CTGAAAGTTT CTTCTTTTTT TCTCTTACTA TATTTATTGA TTTCAAGTCT TCTATTAAGG  | 9681 |
| ACATTTAGCC TTCAATGGAA ATTAAACTC ATTTAGGACT GTATTTCCAA ATTACTGATA   | 9741 |

68

TCAGAGTTAT TTAAAAATTG TTTATTTGAG GAGATAACAT TTCAACTTTG TTCCTAAATA 9801  
 TATAATAATA AAATGATTGA CTTTATTTGC ATTTTATGA CCACTTGTCA TTTATTTTGT 9861  
 CTCGTAAAT TATTTTCATT ATATCAAATA TTTAGTATG TACTTAATAA AATAGGAGAA 9921  
 CATTTTAGAG TTTCAAATTC CCAGGTATTT TCCTTGTTTA TTACCCCTAA ATCATTCCTA 9981  
 TTTAATTCTT CTTTTTAAAT GGAGAAAATT ATGTCTTTTT AATATGGTTT TTGTTTTGTT 10041  
 ATATATTCAC AGGCTGGAGA CGTTTAAAAG ACCGTTTCAA AAGAGATTTA CTTTTTTAAA 10101  
 GGACTTTATC TGAACAGAGA GATATAATAT TTTTCCTATT GGACAATGGA CTGCAAAGC 10161  
 TTCACTTCAT TTTAAGAGCA AAAGACCCCA TGTTGAAAAC TCCATAACAG TTTTATGCTG 10221  
 ATGATAATTT ATCTACATGC ATTTCAATAA ACCTTTTGTT TCCTAAGACT AGATACATGG 10281  
 TACCTTTATT GACCATTAAA AAACCACCAC TTTTGGCAA TTTACCAATT ACAATTGGGC 10341  
 AACCATCAGT AGTAATTGAG TCCTCATTTT ATGCTAAATG TTATGCCTAA CTCTTTGGGA 10401  
 GTTACAAAGG AAATAGCAAT TATGGCTTTT GCCCTCTAGG AGATACAGGA CAAATACAGG 10461  
 AAAATACAGC AACCCAAACT GACAATACTC TATACAAGAA CATAATCACT AAGCAGGAGT 10521  
 CACAGCCACA CAACCAAGAT GCATAGTATC CAAAGTGCAG CTG 10564

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 453 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Met Ser Trp Ser Leu His Pro Arg Asn Leu Ile Leu Tyr Phe Tyr Ala  
 1 5 10 15  
 Leu Leu Phe Leu Ser Ser Thr Cys Val Ala Tyr Val Ala Thr Arg Asp  
 20 25 30

69

Glu Lys Ile His Leu Ile Ser Thr Gln Ser Ala Ile Pro Tyr Ala Leu  
260 265 270

Arg Val Glu Leu Glu Asp Trp Asn Gly Arg Thr Ser Thr Ala Asp Tyr  
 275 280 285  
 Ala Met Phe Lys Val Gly Pro Glu Ala Asp Lys Tyr Arg Leu Thr Tyr  
 290 295 300  
 Ala Tyr Phe Ala Gly Gly Asp Ala Gly Asp Ala Phe Asp Gly Phe Asp  
 305 310 315 320  
 Phe Gly Asp Asp Pro Ser Asp Lys Phe Phe Thr Ser His Asn Gly Met  
 325 330 335  
 Gln Phe Ser Thr Trp Asp Asn Asp Asn Asp Lys Phe Glu Gly Asn Cys  
 340 345 350  
 Ala Glu Gln Asp Gly Ser Gly Trp Trp Met Asn Lys Cys His Ala Gly  
 355 360 365  
 His Leu Asn Gly Val Tyr Tyr Gln Gly Gly Thr Tyr Ser Lys Ala Ser  
 370 375 380  
 Thr Pro Asn Gly Tyr Asp Asn Gly Ile Ile Trp Ala Thr Trp Lys Thr  
 385 390 395 400  
 Arg Trp Tyr Ser Met Lys Lys Thr Thr Met Lys Ile Ile Pro Phe Asn  
 405 410 415  
 Arg Leu Thr Ile Gly Glu Gly Gln Gln His His Leu Gly Gly Ala Lys  
 420 425 430  
 Gln Val Arg Pro Glu His Pro Ala Glu Thr Glu Tyr Asp Ser Leu Tyr  
 435 440 445  
 Pro Glu Asp Asp Leu  
 450

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10807 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

11

(vii) IMMEDIATE SOURCE:

(B) CLONE: ovine beta-lactoglobulin

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

|                                                                    |      |
|--------------------------------------------------------------------|------|
| ACGCGTGTCTG ACCTGCAGGT CAACGGATCT CTGTGTCTGT TTTCATGTTA GTACCACACT | 60   |
| GTTTTGGTGG CTGTAGCTTT CAGCTACAGT CTGAAGTCAT AAAGCCTGGT ACCTCCAGCT  | 120  |
| CTGTTCTCTC TCAAGATTGT GTTCTGCTGT TTGGGTCTTT AGTGTCTCCA CACAATTTTT  | 180  |
| AGAATTGTTT GTTCTAGTTC TGTGAAAAAT GATGCTGGTA TTTTGATAAG GATTGCATTG  | 240  |
| AATCTGTAAA GCTACAGATA TAGTCATTGG GTAGTACAGT CACTTTAACA ATATTAAGTC  | 300  |
| TTCACATCTG TGAGCATGAT ATATTTTCCC CCTCTATATC ATCTTCAATT CCTCCTATCA  | 360  |
| GTTTCTTTCA TTGCAGTTTT CTGAGTACAG GTCTTACACC TCCTTGGTTA GAGTCATTCC  | 420  |
| TCAGTATTTT ATTCCTTTGA TACAATTGTG AATGAGGTAA TTTTCTTAGT TTCTCTTTCT  | 480  |
| GATAGCTCAT TGTTAGTGTA TATATAGAAA AGCAACAGAT TTCTATGTAT TAATTTTGTA  | 540  |
| TCCTGCAACA GATTTCTATG TATTAATTTT GTATCCTGCT ACTTTACGGA ATTCACTTAT  | 600  |
| TAGCTTTTTG GTGACATCTT GAGGATTTTC TGAAGAAAAT GGCATGGTAT GGTAGGACAA  | 660  |
| GGTGTCTATG CATCTGCAAA CAGTGGCAGT TTTCTTCTT CCCTTCCAAC CTGGATTTCT   | 720  |
| TTGATTTCTT TCTGTCTGAG TACGACTAGG ATTCCAATA CTATACCGAA TAAAAGTGGC   | 780  |
| AAGAGTGGAC ATCCTTGTCT TATTTTTCTG ACCTTAGAGG AAATGCTTTC AGTTTTTCAC  | 840  |
| CATTAATTAT AATGTTTACT GTGGGCTTGT CATATGTGGC CTTCAATTATA TGGAGGTCTA | 900  |
| TTCCCTCTAT ACCCACCTTG TTGAGAGTTT TTATCATAAA AGTATGTTGA ATTTTGTCAA  | 960  |
| AAGTTTTTCC TGCATCTATT GAGATGATTT TTAATCTTCA ATTCATTAAT GATTTTTATT  | 1020 |
| CTTCATTTTG TTAATGATTT CCATTCTTCA ATTTGTAAAC GTGGTATATC ACATTGATTG  | 1080 |
| ATTTGTGGAT ACCTTTGTAT CCCTGGGATA AACCTCACTT GATCATGAGC TTTCAATGTA  | 1140 |

|                                                                   |      |
|-------------------------------------------------------------------|------|
| TTTTTGAATT CACTTTGCTA ATATTCTGTT GGGTATTTTT GCATCTCTAT TCATCAATGA | 1200 |
| TATTGGCCTA AGAAAGGTTT TGTCTGGTTT TAGTATCAGG GTGATGCTGG CCTCATAGAG | 1260 |
| AGAGTTTAGA AGCATTTCTT CCTCTTTGAT TTTTCGGAAT AGTTTGAGTA GGATAGGTAT | 1320 |
| TAACTCTTCT TTAAATGTTT GGGGACTTCC CTGGTGAGCC GGTGGTTGAG AATCCGCCTC | 1380 |
| AGGGATGTGG GTTTGATCCC TGGTCAGGGA ACCATTAATA AGATCCCACA TGCTGCAGGC | 1440 |
| AACAAGCCCC CAAGCTGCAA CCACTGAGCT GCAACCGCTG CAGTGCCCAC AGGCCACGAC | 1500 |
| CAGAGAAAGC CCACATACAG CAGGGAAGAC CCAGCACAAC CGGAAAAAGG AGTTTGGTGG | 1560 |
| AATACAGCTG TGAAGCCGTC TGGTCCTGGA CTCCTGCTTG AGGGAATTTT TAAAAATTA  | 1620 |
| TTGATTCAAT TTCATTACTG GTAACGGTC TGTTCATATT TTCTATTTCT TCCGGGTTCA  | 1680 |
| GTCTTGGGAG ATTGTACATG CCTAGGAATG TGTCCGTTTC TTCTAGGTTG TCCATTTTAT | 1740 |
| TGGACATGCA TGGGAGCACA CAGCACCGAC CAGCGAGACT CATGCTGGCT TCCTGGGGCC | 1800 |
| AGGCTGGGGC CCAAGCAGC ATGGCATCCT AGAGTGTGTG AAAGCCCACT GACCCTGCCC  | 1860 |
| AGCCCCACAA TTTCATTCTG AGAAGTGATT CTTGCTTCT GCACTTACAG GCCCAGGATC  | 1920 |
| TGACCTGCTT CTGAGGAGCA GGGGTTTTGG CAGGACGGGG AGATGCTGAG AGCCGACGGG | 1980 |
| GGTCCAGGTC CCCTCCCAGG CCCCCCTGTC TGGGGCAGCC CTTGGGAAAG ATTGCCCCAG | 2040 |
| TCTCCCTCCT ACAGTGGTCA GTCCCAGCTG CCCCAGGCCA GAGCTGCTTT ATTTCCGTCT | 2100 |
| CTCTCTCTGG ATGGTATTCT CTGGAAGCTG AAGGTTCTTG AAGTTATGAA TAGCTTTGCC | 2160 |
| CTGAAGGGCA TGGTTTGTGG TCACGGTTCA CAGGAACTTG GGAGACCCTG CAGCTCAGAC | 2220 |
| GTCCCGAGAT TGGTGGCACC CAGATTTCCT AAGCTCGCTG GGGAACAGGG CGCTTGTTTC | 2280 |
| TCCCTGGCTG ACCTCCCTCC TCCCTGCATC ACCCAGTTCT GAAAGCAGAG CGGTGCTGGG | 2340 |
| GTCACAGCCT CTCGCATCTA ACGCCGGTGT CCAAACCACC CGTGCTGGTG TTCGGGGGGC | 2400 |
| TACCTATGGG GAAGGGCTTC TCACTGCAGT GGTGCCCCC GTCCCTCTG AGATCAGAAG   | 2460 |
| TCCCAGTCCG GACGTCAAAC AGGCCGAGCT CCCTCCAGAG GCTCCAGGGA GGGATCCTTG | 2520 |

|                                                                   |      |
|-------------------------------------------------------------------|------|
| CCCCCCGCT GCTGCCTCCA GTCCTGGTG CCGCACCTT GAGCCTGATC TTGTAGACGC    | 2580 |
| CTCAGTCTAG TCTCTGCCTC CGTGTTTACA CGCCTTCTCC CCATGTCCCC TCCGTGTCCC | 2640 |
| CGTTTTCTCT CACAAGGACA CCGGACATTA GATTAGCCCC TGTTCCAGCC TCACCTGAAC | 2700 |
| AGCTCACATC TGTAAGACC TAGATTCCAA ACAAGATTCC AACCTGAAGT TCCCGGTGGA  | 2760 |
| TGTGAGTTCT GGGGCGACAT CCTTCAACCC CATCAGAGCT TGCAGTTCAT CGCAAAACAT | 2820 |
| GGAACCTGGG GTTTATCGTA AAACCCAGGT TCTTCATGAA AACTGAGCT TCGAGGCTTG  | 2880 |
| TTGCAAGAAT TAAAGGTGCT AATACAGATC AGGGCAAGGA CTGAAGCTGG CTAAGCCTCC | 2940 |
| TCTTTCCATC ACAGGAAAGG GGGGCTGGG GGCGGCTGGA GGTCTGCTCC CGTGAGTGAG  | 3000 |
| CTCTTTCCTG CTACAGTCAC CAACAGTCTC TCTGGGAAGG AAACCAGAGG CCAGAGAGCA | 3060 |
| AGCCGGAGCT AGTTTAGGAG ACCCCTGAAC CTCCACCCAA GATGCTGACC AGCCAGCGGG | 3120 |
| CCCCCTGGAA AGACCCTACA GTTCAGGGGG GAAGAGGGGC TGACCCGCCA GGTCCCTGCT | 3180 |
| ATCAGGAGAC ATCCCCGCTA TCAGGAGATT CCCCCACCTT GCTCCCGTTC CCCTATCCCA | 3240 |
| ATACGCCAC CCCACCCCTG TGATGAGCAG TTTAGTCACT TAGAATGTCA ACTGAAGGCT  | 3300 |
| TTTGCATCCC CTTTGCCAGA GGCACAAGGC ACCCACAGCC TGCTGGGTAC CGACGCCCAT | 3360 |
| GTGGATTGAG CCAGGAGGCC TGTCCTGCAC CCTCCCTGCT CGGGCCCCCT CTGTGCTCAG | 3420 |
| CAACACACCC AGCACCAGCA TTCCCGCTGC TCCTGAGGTC TGCAGGCAGC TCGCTGTAGC | 3480 |
| CTGAGCGGTG TGGAGGGAAG TGTCTGGGA GATTTAAAT GTGAGAGGCG GGAGGTGGGA   | 3540 |
| GGTTGGGCCC TGTGGGCTG CCCATCCAC GTGCCTGCAT TAGCCCCAGT GCTGCTCAGC   | 3600 |
| CGTGCCCCCG CCGCAGGGGT CAGGTCACTT TCCCGTCTG GGGTTATTAT GACTCTTGT   | 3660 |
| ATTGCCATTG CCATTTTGC TACCCTAACT GGGCAGCAGG TGCTTGACAG GCCCTCGATA  | 3720 |
| CCGACCAGGT CCTCCCTCGG AGCTCGACCT GAACCCCATG TCACCCTTGC CCCAGCCTGC | 3780 |
| AGAGGGTGGG TGAAGTGCAG GATCCCTTCA CCAAGGCCA CGGTACATG GTTTGGAGGA   | 3840 |



|                                                                   |      |
|-------------------------------------------------------------------|------|
| GCTGGTGCCC AAGGCAGAGG CCACCCTCCA GGACACACCT GTCCCCAGTG CTGGCTCTGA | 3900 |
| CCTGTCCTTG TCTAAGAGGC TGACCCCGGA AGTGTTCTTG GCACTGGCAG CCAGCCTGGA | 3960 |
| CCCAGAGTCC AGACACCCAC CTGTGCCCCC GCTTCTGGGG TCTACCAGGA ACCGTCTAGG | 4020 |
| CCCAGAGGGG ACTTCCTGCT TGGCCTTGGA TGGAAGAAGG CCTCCTATTG TCCTCGTAGA | 4080 |
| GGAAGCCACC CCGGGGCTG AGGATGAGCC AAGTGGGATT CCGGGAACCG CGTGGCTGGG  | 4140 |
| GGCCCAGCCC GGGCTGGCTG GCCTGCATGC CTCCTGTATA AGGCCCCAAG CCTGCTGTCT | 4200 |
| CAGCCCTCCA CTCCCTGCAG AGCTCAGAAG CACGACCCCA GGGATATCCC TGCAGCCATG | 4260 |
| AAGTGCCTCC TGCTTGCCCT GGGCCTGGCC CTCGCCTGTG GCGTCCAGGC CATCATCGTC | 4320 |
| ACCCAGACCA TGAAAGGCCT GGACATCCAG AAGGTTGAGG GGTGGCCGG GTGGGTGAGT  | 4380 |
| TGCAGGGCGG GCAGGGGAGC TGGGCCTCAG AGAGCCAAGA GAGGCTGTGA CGTTGGGTTC | 4440 |
| CCATCAGTCA GCTAGGGCCA CCTGACAAAT CCCCCTGGG GCAGCTTCAA CCAGGCGTTC  | 4500 |
| ACTGTCTTGC ATTCTGGAGG CTGGAAGCCC AAGATCCAGG TGTGGCAGG GCTGGCTTCT  | 4560 |
| CCTGCGGCCG CTCTCTGGGG AGCAGACGGC CGTCTTCTCC AGTCCTCTGC GCGCCCTGAT | 4620 |
| TTCCTCTTCC TGTGAGGCCA CCAGGCCTGC TGGAACACG CCTGCCTGCG CAGCTTCACA  | 4680 |
| CGACCTTTGT CATCTCTTTA AAGGCCATGT CTCCAGAGTC ATGTGTTGAA GTTCTGGGGG | 4740 |
| TTAGTGGGAC ACAGTTCAGC CCCTAAAAGA GTCTCTCTGC CCCTCAAATT TTCCCCACCT | 4800 |
| CCAGCCATGT CTCCCCAAGA TCCAAATGTT GCTACATGTG GGGGGGCTCA TCTGGGTCCC | 4860 |
| TCTTTGGGTT CAGTGTGAGT CTGGGGAGAG CATTCCCCAG GGTGCAGAGT TGGGGGGAGT | 4920 |
| ATCTCAGGGC TGCCCAGGCC GGGGTGGGAC AGAGAGCCCA CTGTGGGGCT GGGGGCCCCT | 4980 |
| TCCCACCCCC AGAGTGCAAC TCAAGGTCCC TCTCCAGGTG GCGGGGACTT GGCACCTCTT | 5040 |
| GGCTATGGCG GCCAGCGACA TCTCCCTGCT GGATGCCCAG AGTGCCCCC TGAGAGTGTA  | 5100 |
| CGTGGAGGAG CTGAAGCCCA CCCCCGAGG CAACCTGGAG ATCCTGCTGC AGAAATGGTG  | 5160 |
| GGCGTCTCTC CCAACATGG AACCCCCACT CCCAGGGCT GTGGACCCCC CGGGGGGTGG   | 5220 |

|                                                                   |      |
|-------------------------------------------------------------------|------|
| GGTGCAGGAG GGACCAGGGC CCCAGGGCTG GGGAAGAGGG CTCAGAGTTT ACTGGTACCC | 5280 |
| GGCGCTCCAC CCAAGGCTGC CCACCCAGGG CTTTTTTTTT TTTTAACTT TTATTAATTT  | 5340 |
| GATGCTTCAG AACATCATCA AACAAATGAA CATAAAACAT TCATTTTTGT TTACTTGGA  | 5400 |
| GGGGAGATAA AATCCTCTGA AGTGGAAATG CATAGCAAAG ATACATACAA TGAGGCAGGT | 5460 |
| ATTCTGAATT CCCTGTTAGT CTGAGGATTA CAAGTGTATT TGAGCAACAG AGAGACATTT | 5520 |
| TCATCATTTT TAGTCTGAAC ACCTCAGTAT CTAAAATGAA CAAGAAGTCC TGGAAACGAA | 5580 |
| GCAGTGTGGG GATAGGCCCG TGTGAAGGCT GCTGGGAGGC AGCAGACCTG GGTCTTCGGG | 5640 |
| CTCAAGCAGT TCCCCTACC AGCCCTGTCC ACCTCAGACG GGGGTCAGGG TGCAGGAGAG  | 5700 |
| AGCTGGATGG GTGTGGGGG AGAGATGGGG ACCTGAACCC CAGGGCTGCC TTTTGGGGT   | 5760 |
| GCCTGTGGTC AAGGCTCTCC CTGACCTTTT CTCTCTGGCT TCATCTGACT TCTCCTGGCC | 5820 |
| CATCCACCCG GTCCCCTGTG GCCTGAGGTG ACAGTGAGTG CGCCGAGGCT AGTTGGCCAG | 5880 |
| CTGGCTCCTA TGCCCATGCC ACCCCCCTCC AGCCCTCCTG GGCCAGCTTC TGCCCTGGC  | 5940 |
| CCTCAGTTCA TCCTGATGAA AATGGTCCAT GCCAATGGCT CAGAAAGCAG CTGTCTTTCA | 6000 |
| GGGAGAACGG CGAGTGTGCT CAGAAGAAGA TTATTGCAGA AAAAACCAAG ATCCCTGCGG | 6060 |
| TGTTCAAGAT CGATGGTGAG TCCGGGTCCC TGGGGGACAC CCACCACCCC CGCCCCGGG  | 6120 |
| GACTGTGGAC AGGTTCAAGG GGCTGGCGTC GGGCCCTGGG ATGCTAAGGG ACTGGTGGTG | 6180 |
| ATGAAGACAC TGCCTTGACA CCTGCTTCAC TTGCCTCCCC TGCCACCTGC CCGGGGCCTT | 6240 |
| GGGGCGGTGG CCATGGGCAG GTCCCGGCTG GCGGGCTAAC CCACCAGGGT GACACCCGAG | 6300 |
| CTCTCTTTGC TGGGGGGCGG GCGGTGCTCT GGGCCCTCAG GCTGAGCTCA GGAGGTACCT | 6360 |
| GTGCCCTCCC AGGGGTAACC GAGAGCCGTT GCCCACTCCA GGGGCCAGG TGCCCCACGA  | 6420 |
| CCCCAGCCCG CTCCACAGCT CCTTCATCTC CTGGAGACAA ACTCTGTCCG CCCTCGCTCA | 6480 |
| TTCACTTGTT CGTCCTAAAT CCGAGATGAT AAAGCTTCGA GGGGGGGTTG GGGTTCCATC | 6540 |

|                                                                   |      |
|-------------------------------------------------------------------|------|
| AGGGCTGCCC TTCCGCCGGG CAGCCTGGGC CACATCTGCC CTTGGCCCCC TCAGGACTCA | 6600 |
| CTCTGACTGG AGGCCCTGCA CTGACTGACG CCAGGGTGCC CAGCCCAGGG TCTCTGGCGC | 6660 |
| CATCCAGCTG CACTGGGTTT GGGTGCTGGT CCTGCCCCCA AGCTGCCCCG ACACCACAGG | 6720 |
| CAGCCGGGGC TGCCCACTGG CCTCGGTCAG GGTGAGCCCC AGCTGCCCCC GCTCAGGGCT | 6780 |
| TGCCCCGACA ATGACCCCAT CCTCAGGACG CACCCCCCTT CCCTTGCTGG GCAGTGTTCA | 6840 |
| GCCCCACCCG AGATCGGGGG AAGCCCTATT TCTTGACAAC TCCAGTCCCT GGGGGAGGGG | 6900 |
| GCCTCAGACT GAGTGGTGAG TGTTCCCAAG TCCAGGAGGT GGTGGAGGGT CCTGGCGGAT | 6960 |
| CCAGAGTTGA CAGTGAGGGC TTCCTGGGCC CCATGCGCCT GGCAGTGGCA GCAGGGAAGA | 7020 |
| GGAAGCACCA TTTCAGGGGT GGGGGATGCC AGAGGCGCTC CCCACCCCGT CTTGCCGGG  | 7080 |
| TGGTGACCCC GGGGGAGCCC CGCTGGTCGT GGAGGGTGCT GGGGGCTGAC TAGCAACCCC | 7140 |
| TCCCCCCCCG TTGGAActCA CTTTTCTCCC GTCTTGACCG CGTCCAGCCT TGAATGAGAA | 7200 |
| CAAAGTCCTT GTGCTGGACA CCGACTACAA AAAGTACCTG CTCTTCTGCA TGGAAAACAG | 7260 |
| TGCTGAGCCC GAGCAAAGCC TGGCCTGCCA GTGCCTGGGT GGGTGCCAAC CCTGGCTGCC | 7320 |
| CAGGGAGACC AGCTGCGTGG TCCTTGCTGC AACAGGGGGT GGGGGGTGGG AGCTTGATCC | 7380 |
| CCAGGAGGAG GAGGGGTGGG GGGTCCCTGA GTCCCGCCAG GAGAGAGTGG TCGCATACCG | 7440 |
| GGAGCCAGTC TGCTGTGGGC CTGTGGGTGG CTGGGGACGG GGGCCAGACA CACAGGCCGG | 7500 |
| GAGACGGGTG GGCTGCAGAA CTGTGACTGG TGTGACCGTC GCGATGGGGC CGGTGGTCAC | 7560 |
| TGAATCTAAC AGCCTTTGTT ACCGGGGAGT TTCAATTATT TCCAAAATA AGAACTCAGG  | 7620 |
| TACAAAGCCA TCTTTCAACT ATCACATCCT GAAAACAAAT GGCAGGTGAC ATTTTCTGTG | 7680 |
| CCGTAGCAGT CCCACTGGGC ATTTTCAGGG CCCCTGTGCC AGGGGGGCGC GGGCATCGGC | 7740 |
| GAGTGGAGGC TCCTGGCTGT GTCAGCCGGC CCAGGGGGAG GAAGGGACCC GGACAGCCAG | 7800 |
| AGGTGGGGGG CAGGCTTTCC CCCTGTGACC TGCAGACCCA CTGCACTGCC CTGGGAGGAA | 7860 |
| GGGAGGGGAA CTAGGCCAAG GGGGAAGGGC AGGTGCTCTG GAGGGCAAGG GCAGACCTGC | 7920 |

|            |             |             |             |            |            |      |
|------------|-------------|-------------|-------------|------------|------------|------|
| AGACCACCCT | GGGGAGCAGG  | GACTGACCCC  | CGTCCCTGCC  | CCATAGTCAG | GACCCCGGAG | 7980 |
| GTGGACAACG | AGGCCCTGGA  | GAAATTCGAC  | AAAGCCCTCA  | AGGCCCTGCC | CATGCACATC | 8040 |
| CGGCTTGCTT | TCAACCCGAC  | CCAGCTGGAG  | GGTGAGCACC  | CAGGCCCCGC | CCTTCCCCAG | 8100 |
| GGCAGGAGCC | ACCCGGCCCC  | GGGACGACCT  | CCTCCCATGG  | TGACCCCCAG | CTCCCCAGGC | 8160 |
| CTCCCAGGAG | GAAGGGGTGG  | GGTGCAGCAC  | CCCGTGGGGG  | CCCCCTCCCC | ACCCCTGCC  | 8220 |
| AGGCCTCTCT | TCCCGAGGTG  | TCCAGTCCCA  | TCCTGACCCC  | CCCATGACTC | TCCCTCCCCC | 8280 |
| ACAGGGCAGT | GCCACGTCTA  | GGTGAGCCCC  | TGCCGGTGCC  | TCTGGGGTAA | GCTGCCTGCC | 8340 |
| CTGCCCCACG | TCCTGGGCAC  | ACACATGGGG  | TAGGGGTCT   | TGGTGGGGCC | TGGGACCCCA | 8400 |
| CATCAGGCCC | TGGGGTCCCC  | CCTGTGAGAA  | TGGCTGGAAG  | CTGGGGTCCC | TCCTGGCGAC | 8460 |
| TGCAGAGCTG | GCTGGCCGCG  | TGCCACTCTT  | GTGGGTGACC  | TGTGTCCTGG | CCTCACACAC | 8520 |
| TGACCTCCTC | CAGCTCCTTC  | CAGCAGAGCT  | AAGGCTAAGT  | GAGCCAGAAT | GGTACCTAAG | 8580 |
| GGGAGGCTAG | CGGTCCTTCT  | CCCGAGGAGG  | GGCTGTCCTG  | GAACCACCAG | CCATGGAGAG | 8640 |
| GCTGGCAAGG | GTCTGGCAGG  | TGCCCCAGGA  | ATCACAGGGG  | GGCCCCATGT | CCATTTCAGG | 8700 |
| GCCCCGGAGC | CTTGGA CTCC | TCTGGGGACA  | GACGACGTCA  | CCACCGCCCC | CCCCCATCA  | 8760 |
| GGGGGACTAG | AAGGGACCAG  | GACTGCAGTC  | ACCCTTCCTG  | GGACCCAGGC | CCCTCCAGGC | 8820 |
| CCCTCCTGGG | GCTCCTGCTC  | TGGGCAGCTT  | CTCCTTCACC  | AATAAAGGCA | TAAACCTGTG | 8880 |
| CTCTCCCTTC | TGAGTCTTTG  | CTGGACGACG  | GGCAGGGGGT  | GGAGAAGTGG | TGGGGAGGGA | 8940 |
| GTCTGGCTCA | GAGGATGACA  | GCGGGGCTGG  | GATCCAGGGC  | GTCTGCATCA | CAGTCTTGTG | 9000 |
| ACAACTGGGG | GCCCACACAC  | ATCACTGCGG  | CTCTTTGAAA  | CTTTCAGGAA | CCAGGGAGGG | 9060 |
| ACTCGGCAGA | GACATCTGCC  | AGTTCAC TTG | GAGTGTT CAG | TCAACACCCA | AACTCGACAA | 9120 |
| AGGACAGAAA | GTGGAAAATG  | GCTGTCTCTT  | AGTCTAATAA  | ATATTGATAT | GAAACTCAAG | 9180 |
| TTGCTCATGG | ATCAATATGC  | CTTTATGATC  | CAGCCAGCCA  | CTACTGTCGT | ATCAACTCAT | 9240 |

|                                                                    |       |
|--------------------------------------------------------------------|-------|
| GTACCCAAAC GCACTGATCT GTCTGGCTAA TGATGAGAGA TTCCCAGTAG AGAGCTGGCA  | 9300  |
| AGAGGTCACA GTGAGAACTG TCTGCACACA CAGCAGAGTC CACCAGTCAT CCTAAGGAGA  | 9360  |
| TCAGTCCTGG TGTTCATTGG AGGACTGATG TTGAAGCTGA AACTCCAATG CTTTGGCCAC  | 9420  |
| CTGATGTGAA GAGCTGACTC ATTTGAAAAG ACCCTGATGC TGGGAAAGAT TGAGGGCAGG  | 9480  |
| AGGAGAAGGG GACGACAGAG GATGAGATGG TTGGATGGCA TCACCAACAC AATGGACATG  | 9540  |
| GGTTTGGGTG GACTCCAGGA GTTGGTGATG GACAGGGAGG CCTGGCGTGC TACGGAAGCG  | 9600  |
| GTTTATGGGG TCACAAAGAC TGAGTGA CTG AACTGAGCTG AACTGAATGG AAATGAGGTA | 9660  |
| TACAGCAAAG TGGGGATTTT TTAGATAATA AGAATATACA CATAACATAG TGTATACTCA  | 9720  |
| TATTTTTATG CATACTGAA TGCTCAGTCA CTCAGTCGTA TCTGACTCTG TGACCTATGG   | 9780  |
| ACCGTAGCCT TCCAGGTTTC TTCTGTCCAC AGAATTCTCC AAGGCAAGAA TACTGGAGTG  | 9840  |
| GGTAGCCATT TCCTCCTCCA GGGGATCCTC CCGACCCAGG GATTGAACCG GCATCTCCTG  | 9900  |
| TATTGGCAGG TGGATTCTTT ACCACTGTGC CACCAGGGAA GCCCGTGTTA CTCTCTATGT  | 9960  |
| CCCACTTAAT TACCAAAGCT GCTCCAAGAA AAAGCCCCTG TGCCCTCTGA GCTTCCCGGC  | 10020 |
| CTGCAGAGGG TGGTGGGGGT AGACTGTGAC CTGGGAACAC CCTCCCGCTT CAGGACTCCC  | 10080 |
| GGGCCACGTG ACCCACAGTC CTGCAGACAG CCGGGTAGCT CTGCTCTTCA AGGCTCATT   | 10140 |
| TCTTTAAAAA AACTGAGGT CTATTTTGTG ACTTCGCTGC CGTAACTTCT GAACATCCAG   | 10200 |
| TGCGATGGAC AGGACCTCCT CCCCAGGCCT CAGGGGCTTC AGGGAGCCAG CCTTCACCTA  | 10260 |
| TGAGTCACCA GACACTCGGG GGTGGCCCCG CCTTCAGGGT GCTCACAGTC TTCCATCGT   | 10320 |
| CCTGATCAAA GAGCAAGACC AATGACTTCT TAGGAGCAAG CAGACACCCA CAGGACACTG  | 10380 |
| AGGTTACCA GAGCTGAGCT GTCCTTTTGA ACCTAAAGAC ACACAGCTCT CGAAGGTTTT   | 10440 |
| CTCTTTAATC TGGATTTAAG GCCTACTTGC CCCTCAAGAG GGAAGACAGT CCTGCATGTC  | 10500 |

CCCAGGACAG CCACTCGGTG GCATCCGAGG CCACTTAGTA TTATCTGACC GCACCCTGGA 10560  
 ATTAATCGGT CCAAACCTGA CAAAACCTT GGTGGGAAGT TTCATCCCAG AGGCCTCAAC 10620  
 CATCCTGCTT TGACCACCCT GCATCTTTTT TTCTTTTATG TGTATGCATG TATATATATA 10680  
 TATATATTTT TTTTTTTTTT ATTTTTTGGC TGTGCTGGCT GTTCGTTGCA GTTCGGTGCG 10740  
 CAGGCTTCTC TCTAGTTTCT CTCTAGTCTT CTCTTATCAC AGAGCAGTCT CTAGACGATC 10800  
 GACGCGT 10807

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 47 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

AATTCCGATC GACGCGTCGA CGATATACTC TAGACGATCG ACGCGTA 47

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:

- (B) CLONE: BLGAMP3

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

TGGATCCCCT GCCGGTGCCT CTGG 24

## (2) INFORMATION FOR SEQ ID NO:10:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (vii) IMMEDIATE SOURCE:

- (B) CLONE: BLGAMP4

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

AACGCGTCAT CCTCTGTGAG CCAG

24

## (2) INFORMATION FOR SEQ ID NO:11:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (vii) IMMEDIATE SOURCE:

- (B) CLONE: ZC6839

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

ACTACGTAGT

10

## (2) INFORMATION FOR SEQ ID NO:12:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 42 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (vii) IMMEDIATE SOURCE:

- (B) CLONE: ZC6632

81

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

CGACGCGGAT CCTACGTACC TGCAGCCATG TTTTCATGA GG

42

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:

(B) CLONE: ZC6627

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

AGGGCTTCGG CAAGCTTCAG G

21

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:

(B) CLONE: ZC6521

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

GCCAAAGACT TACTTCCTC TAGA

24



## (2) INFORMATION FOR SEQ ID NO:15:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (vii) IMMEDIATE SOURCE:

- (B) CLONE: ZC6520

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

GCATGAACGT CGCGTGGTGG TTGTGCTACC

30

## (2) INFORMATION FOR SEQ ID NO:16:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (vii) IMMEDIATE SOURCE:

- (B) CLONE: ZC6519

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

ACCACGCGAC GTTCATGCTC TAAAACCGTT

30

## (2) INFORMATION FOR SEQ ID NO:17:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 36 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (vii) IMMEDIATE SOURCE:

- (B) CLONE: ZC6518

83

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

GCTGCGGGAT CCTACGTACT AGGGGGACAG GGAAGG

36

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 45 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:

(B) CLONE: ZC6629

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

CGACGCGAAT TCTACGTACC TGCAGCCATG AAAAGGATGG TTTCT

45

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 45 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:

(B) CLONE: ZC6630

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

CGACGCGAAT TCTACGTACC TGCAGCCATG AAACATCTAT TATTG

45

84

## (2) INFORMATION FOR SEQ ID NO:20:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 21 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

- (vii) IMMEDIATE SOURCE:
  - (B) CLONE: ZC6625

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

GTGAGATTTT CAGATCTTGT C

21

## (2) INFORMATION FOR SEQ ID NO:21:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 21 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

- (vii) IMMEDIATE SOURCE:
  - (B) CLONE: ZC6626

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

AAGAATTACT GTGGCCTACC A

21

## (2) INFORMATION FOR SEQ ID NO:22:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 33 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

85

(vii) IMMEDIATE SOURCE:  
(B) CLONE: ZC6624

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

GCTGCGGAAT TCTACGTACT ATTGCTGTGG GAA

33

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 45 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:  
(B) CLONE: ZC6514

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

CGACGCGGAT CCTACGTACC TGCAGCCATG AGTTGGTCCT TGCAC

45

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 21 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:  
(B) CLONE: zc6517

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

GTCTCTGGTA GCAACATACT A

21



## (2) INFORMATION FOR SEQ ID NO:25:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (vii) IMMEDIATE SOURCE:

- (B) CLONE: zc6516

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

GGGTTTCTAG CCCTACTAGT AG

22

## (2) INFORMATION FOR SEQ ID NO:26:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (vii) IMMEDIATE SOURCE:

- (B) CLONE: zc6515

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

GGGTTTCTAG CCCTACTAGT AG

22

(2) INFORMATION FOR SEQ ID NO:27:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 47 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

AAGCTACGCG TCGATCGTCT AGAGTATATC GTCGACGCGT CGATCGG